DOCUMENT RESUME

ED 467 733 FL 027 421

AUTHOR	Cahill, Michael, Ed.
TITLE	Notes on Linguistics, 2001.
INSTITUTION	SIL International, Dallas, TX.
ISSN	ISSN-0736-0673
PUB DATE	2001-00-00
NOTE	265p.; Published quarterly.
AVAILABLE FROM	International Academic Bookstore, 7500 West Camp Wisdom Road, Dallas, TX 75236 ($3 per issue; $20.95 per year; $23.95 per year outside U.S.). Tel: 972-708-7404; e-mail: academic_books@sil.org.
PUB TYPE	Collected Works - Serials (022)
JOURNAL CIT	Notes on Linguistics; v4 n1-4 2001
EDRS PRICE	EDRS Price MF01/PC11 Plus Postage.
DESCRIPTORS	American Indians; Archives; Athapascan Languages; Bantu Languages; Bilingualism; *Contrastive Linguistics; Dictionaries; Foreign Countries; German; Grammar; Language Processing; Lexicography; Mandarin Chinese; Mexicans; Morphology (Languages); Phonology; Second Language Instruction; Semantics; Syntax; Turkish; *Uncommonly Taught Languages; World Wide Web
IDENTIFIERS	Amazonian Languages; Balkan Languages; Bilingual Dictionaries; Central African Republic; Endangered Languages; Papua New Guinea; Southeast Asians

ABSTRACT

These four volumes of the journal present articles, review articles, abstracts, dissertation abstracts, and reports. Articles include the following: "The Linguist's Role in Archiving Linguistic Data Sources" (Joan Spanne); "The SIL Language and Culture Archive: An Interview with Joan Spanne" (Eugene Loos); "The Value of Comparative Linguistics" (Joseph E. Grimes); "Historical Linguistics in Southeast Asian Language Programs" (Paulette Hopple); "Comparative-Historical MesoAmerican Reconstruction and SIL Personnel: Accomplishments and Problems" (Robert E. Longacre); "The Impact of Bilingual Dictionaries in Mexican Indian Languages" (Doris Bartholomew); "Lexicography and Mass Production" (Ronald Moe); "Lexicography in the Field: Methods and Results of the MUNA Dictionary Project" (Rene van den Berg); and "Global Language Viability: Causes, Symptoms and Cures for Endangered Languages" (Barbara F. Grimes). (Papers contain references.) (SM)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 467 733

Notes on Linguistics, 2001.

Volume 4, Numbers 1-4, 2001

Michael Cahill, Editor

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality

Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

E. Loos

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

FL 027 421

2

BEST COPY AVAILABLE

NOTES ON *LINGUISTICS*

VOLUME 4, NUMBER 1 **2001**

CONTENTS

FROM THE LINGUISTICS DEPARTMENT Michael Cahill 3

ARTICLES

THE LINGUIST'S ROLE IN ARCHIVING LINGUISTIC DATA RESOURCES Joan Spanne 4

THE SIL LANGUAGE AND CULTURE ARCHIVE: AN INTERVIEW WITH JOAN SPANNE Eugene Loos 16

REVIEW ARTICLE

THE AMAZONIAN LANGUAGES by R.M.W. Dixon and Alexandra Y. Aikhenvald, eds. Terry Malone 29

DISSERTATION ABSTRACTS

THE INTERFACE BETWEEN SYNTAX AND DISCOURSE IN KORAFE, A PAPUAN LANGUAGE OF PAPUA NEW GUINEA Cynthia J.M. Farr 50

THE DYNAMICS OF LANGUAGE SPREAD: A STUDY OF THE MOTIVATIONS AND THE SOCIAL DETERMINANTS OF THE SPREAD OF SANGO IN THE REPUBLIC OF CENTRAL AFRICA Mark Edward Karan 51

REPORTS

WEB-BASED LANGUAGE DOCUMENTATION AND DESCRIPTION AND THE OPEN LANGUAGE ARCHIVES COMMUNITY WORKSHOP J. Albert Bickford 52

INTERNATIONAL LINGUISTIC ASSOCIATION 46TH ANNUAL CONFERENCE LANGUAGES OF THE AMERICAS, NATIVE AND NON-NATIVE Terry Malone 58

(Continued on back cover)

SIL INTERNATIONAL
7500 W. CAMP WISDOM ROAD
DALLAS, TEXAS 75236 USA

FLOL 1484

NOTES ON LINGUISTICS

INTERNATIONAL LINGUISTICS COORDINATOR/EDITOR ■
Michael Cahill
ASSOCIATE EDITOR ■ *Eugene Loos*
CONSULTING EDITOR ■ *Alan S. Kaye*
FORMAT EDITOR / SECRETARY ■ *Betty Philpott*

INTERNATIONAL LINGUISTICS ADVISORS

Thomas Bearth ..... *U. of Zurich*
Steven Bird ..... *U.of Pennsylvania*
Ruth Brend ..... *Michigan State U., r.*
Derek Fivaz ..... *Rhodes U., r.*
Peter Fries ..... *Central Michigan U.*
Shirley Brice Heath ..... *Stanford U.*
Yoshihiko Ikegami ..... *U. of Tokyo*
Frances Ingemann ..... *U. of Kansas*
Peter Ladefoged ..... *U.C.L.A.*
Sydney Lamb ..... *Rice U.*
W. P. Lehmann ..... *U. of Texas at Austin*
Howard P. McKaughan ..... *U. of Hawaii, r.*
Dinh Hoa Nguyen ..... *Southern Illinois U., r.*
John Oller ..... *U. of Southwestern Louisiana*
A. K. Pawley ..... *Australian National U.*
Richard Rhodes ..... *U. of California, Berkeley*
Malcolm Ross ..... *Australian National U.*
William J. Samarin ..... *U. of Toronto*
Hu T'an ..... *N.R.C. on Tibetology, Beijing*
Rudolph Troike ..... *U. of Arizona*
Richard Tucker ..... *Carnegie Mellen U.*
Robert Van Valin ..... *SUNY, Buffalo*
Ekkehard Wolff ..... *U. of Leipzig*

INTERNATIONAL LINGUISTICS CONSULTANTS

Rudy Barlaan ..... *Philippines*
John Bendor-Samuel ..... *U.K.*
Andy Black ..... *Mexico*
Cheri Black ..... *Mexico*
Regina Blass ..... *Burkina Faso/Niger*
Les Bruce ..... *Dallas*
Donald A. Burquest ..... *U.T., Arlington*
Gene Casad ..... *Mexico*
Rod Casali ..... *Africa*
Ross Caughley ..... *Asia*
Lon Diehl ..... *Asia*
Daniel Everett ..... *Brazil*
Warren Glover ..... *South Asia*
Kenneth J. Gregerson ..... *Asia*
Joseph Grimes ..... *Hawaii*
Austin Hale ..... *South Asia*
Lou Hohulin ..... *Dallas*
George Huttar ..... *Nairobi*
Connie Kutsch Lojenga ..... *Africa*
Stephen H. Levinsohn ..... *Colombia*
Ivan Lowe ..... *U.K.*
Steve Marlett ..... *Mexico*
Mike Maxwell ..... *Waxhaw*
Kenneth McElhanon ..... *Dallas*
Doris Payne ..... *U. of Oregon*
Tom Payne ..... *U. of Oregon*
Gary Simons ..... *Dallas*
Keith Snider ..... *Canada*
David Weber ..... *Peru*
Klaus Wedekind ..... *East Africa*
Ursula Wiesemann ..... *Germany*

SIL International President ■ *Carolyn Miller*
SIL International Executive Director ■ *John Watters*
SIL International VP for Academic Affairs ■ *Karl Franklin*

NOTES FOR READERS ■ Permission is hereby granted by this journal to reproduce material from this issue, if readers obtain similar permission directly from its authors and if credit is given the authors and *Notes on Linguistics*.

NOTES FOR SUBSCRIBERS ■ *Notes on Linguistics* is a quarterly publication of the Linguistics Department of SIL International, designed to share information of practical, theoretical, administrative, or general interest. It is specifically intended to provide linguistic field workers with news, reviews, announcements, and articles that will stimulate interest in linguistics and help them stay current.

SUBSCRIPTION ■ Annual subscription rate: U.S address (delivered by U.S. Postal Service) $20.95; Outside U.S. (delivered by AIR) $23.95. A 20% discount is offered to resellers and personnel of SIL International. Prices include postage and handling. Subscriptions will be prorated so that all subscriptions end with the fourth issue of the volume.

INDIVIDUAL ISSUES ■ Individual copies of available issues are $3.00 each, plus a minimum of $2.75 handling (which generally covers about 3 issues). Not all back issues are still available.

ORDERING INFORMATION: ■ Prepayment is requested from individuals; reseller and institutional purchase orders are accepted. Payment must be in U.S. funds, and checks must be drawn on a U.S. bank. Major credit cards are welcomed; to pay by credit card please include account number and expiration date. Electronic transfers are not accepted. Subscriptions, standing orders, and back issues are handled by: International Academic Bookstore, 7500 W. Camp Wisdom Road, Dallas, TX 75236, USA. Phone: 972-708-7404 FAX: 972-708-7363; Email: academic_books@sil.org

NOTES FOR CONTRIBUTORS ■ Readers are invited to submit publishable materials or letters of comment to the editor of *Notes on Linguistics* at The International Linguistics Center, 7500 W. Camp Wisdom Road, Dallas, TX 75236 USA. Internet e-mail: Mike_Cahill@sil.org (articles) or EugeneLoos@compuserve.com (reviews). Intending contributors should ensure that their manuscripts conform to the standards laid down in the LSA Style Sheet printed periodically in the *LSA Bulletin*. Contributors who have copies of their manuscripts on magnetic media such as diskettes in PC/MS-DOS are urged to submit the same along with their manuscript. (3½" PC/MS-DOS diskettes or as e-mail attachments, in Word for Windows format are preferred.)

DISCLAIMER: ■ Responsibility for the accuracy of quoted and referenced material rests solely upon the author of the item in which such material appears. Opinions expressed by the authors do not necessarily reflect policies of SIL.

ISSN 0736-0673

BEST COPY AVAILABLE

From the Linguistics Coordinator

How can *Notes on Linguistics* best meet the needs of the field linguist? That is the question we continually want to keep in mind as we assemble each issue. Different people will answer this in different ways, of course, and I am quite open to ideas about how to improve this publication.

For the next few issues, and possibly beyond, I would like to try something a little different than usual, and that is to have 'thematic issues'. These issues will have two or three articles which focus on one topic, one relevant and useful to field workers. In future issues, I expect to see articles dealing with dictionaries, with experimental linguistics (even without special equipment!), with the uses of historical and comparative linguistics, with computer tools, and so on. Suggestions for future themes are most welcome.

In this issue, we focus on the topic of archiving language data. As notebooks get older, as papers in the boxes under your beds age and turn yellow, as tapes crack and degrade, as you've lost the copy of the computer program you used in the 1980s, the problem of preserving our language data becomes more and more acute. It is crucial to have strategies in place to preserve the language data we have worked so hard to collect.

So here we present both an article and an interview with Joan Spanne, who is in charge of SIL's Language and Culture Archives. She discusses the urgency and the challenges of archiving data (it's not the same as backing up your computer files) as well as some 'how-to' tips. We also include a report by Albert Bickford on an interesting conference on internet archiving, and a proposal to create a directory through which all language materials on the internet can be located. I hope you find these interesting and useful.

One other item I would like to include more of is recognition of significant linguistic achievements by SIL members. We have traditionally included dissertation abstracts in *NOLx*, but some grammars and dictionaries involve just as much work, and are also worthy of recognition. If you know of such a work recently published by an SIL member, please let us know.

Finally, I would like in this first issue of *NOLx* for which I am editor, to gratefully acknowledge the help of Eugene Loos and Betty Philpott. They do most of the work, and I am thankful to have them around.

Michael Cahill
International Linguistics Coordinator

■

The linguist's role in archiving linguistic data resources

Joan Spanne
Director of SIL Language and Culture Archives

1. Introduction. 'What do I need to do to archive my linguistic materials?' Several converging factors are leading more and more linguists to seek answers to this complex question. Some of these factors are:

- The rate of language change (and language death), which heightens the urgency of preserving information about minority and endangered languages (a supply factor—we must not lose what evidence we have).

- The awareness that the language community, as well as individual speakers of a language of study, have an interest in the disposition (preservation, access, use) of research materials about their linguistic and cultural heritage (an ethical factor).

- The expanding availability of sophisticated analytical tools in the linguist's arsenal, which enable the researcher to apply diverse methods for study and analysis to source texts and data sets (a demand factor).

- The mounting evidence that data resources of the computer era are relatively fragile and short-lived—easily corrupted or made obsolete by advancing technologies which are not fully backward-compatible (a time factor: the critical time frame for instigating preservation strategies is much shorter than in the pre-computer era) (Rothenberg 1999; Bearman 1999).

Certainly other pressures for archiving also exist, depending on the circumstances of the researcher. As a result of these pressures, the linguistic research community is beginning to work together to formulate standards and best practices for resource description, preservation, access, and tool development which will best serve the needs of all the interested parties: researchers, language communities, software developers, archival repositories (and their supporting institutions) and even businesses with interests in linguistic computing. The Open Language Archives Community is one significant forum in which this development is taking place.[1] OLAC and its participating archival repositories and research institutes are working

[1] The Open Language Archives Community and the December 2000 workshop 'Web-Based Language Documentation and Description' which launched it are the subject of another paper in this journal issue and will not be described in depth here.

on answers to the complex questions involved in archiving all types of language research materials and enabling access and use through Internet-based repositories, taking advantage of the vastly increased capacity for sharing complex information resources which the growth of the Internet affords. This article is intended to aid linguists and anthropologists in preparing their primary source materials for deposit with an archival repository, so that these valuable resources can be preserved for the long term and made accessible to other users.

2. What should be archived, and what formats are best? Archiving everything might be a goal in an ideal world, but in the real world of limited resources for storage and management, and limited time for all the work of describing, maintaining, finding and using materials, a certain amount of selectivity in archival work is necessary. The researcher preparing materials for deposit in some repository is the first-line selector, separating the wheat from the chaff among the materials he/she has collected and developed in the research process. Table 1 gives a list of the more prominent types of language resources desirable for archiving, though the list is not intended to be comprehensive. This article focuses on archiving primary source materials: recordings and transcriptions, lexical data, word lists, original texts, and field notes.

Table 1: Resource types and formats

Type of material	Format sought for archiving:
Language text, e.g., transcription of recording, original written text, translated text	Paper print-out from a formatted document that includes fonts correctly rendered; Standard Format or tagged (XML) text
Word list	Paper print-out from a formatted document that includes fonts correctly rendered; Standard Format or tagged (XML) text
Sound recording	Magnetic tape recording (cassette or reel-to-reel); WAV file
Lexical or anthropological data file	SFM or XML tagged file with supporting settings files and descriptive documentation
Descriptive or analytical document, e.g., working paper, article, report	Paper print out from a formatted document that includes fonts correctly rendered; HTML, Rich Text Format, Portable Document Format
Font and character set rendered by it	Font files with complete printed description of code points and characters rendered

For all textual materials, the documentation of character encoding is critical; be sure that any specialized fonts needed to render the data are also included and documented. A paper print-out is a very important resource even for materials to be used in digital format in computer-based analysis, as it provides an accuracy check for rendering the computer data.

The value of untranscribed recordings from unknown (undocumented) speech events or speakers is relatively low, as so many factors work against another researcher being able to make use of them. Questionable transcriptions (e.g. those made using inconsistent conventions or before reasonable familiarity with the language was developed) are also of relatively low value, especially if the source recording is lost. A significant factor in favor of preserving such materials would be the relative rarity of materials in the language—the fewer the number of resources in or about a particular language, the more valuable those few resources are. In such a case, additional work on such materials by someone quite knowledgeable in the language probably will be necessary.

Descriptive and analytical works in long-outdated proprietary computer formats are usually best preserved in print format, as the accurate recovery or conversion and maintenance of such works in computer format is very difficult. For these works, the value lies in their intellectual content, rather than in the capability to use the file in some computer-based processing. The archival repository will be responsible for determining the most suitable way of making print materials available to its users (perhaps through photocopying or scanning).

Tape recordings on magnetic media require special work to preserve them and careful procedures to reformat them digitally. This work should be carried out by the repository or a preservation specialist within strictly controlled parameters.[2] For these, and any materials deposited with an archival repository, the depositor may request copies of reformatted works (e.g. a copy of a digitized sound recording made from an original magnetic recording) if he or she desires.

3. Describing linguistic research materials. In order to be able to find a resource in an archival repository, to manage access to the resource, to preserve it through generations of technological change and to know how to use it once obtained, the resource needs to be described in precise ways. The essential work for the linguist is to understand and provide the information needed as much as is realistically possible. The rest of this article discusses

[2] Documenting the characteristics of digital audio files for archival purposes is a subject requiring more expertise than I can claim and more space than is available here.

the common elements used in describing resources, and suggests some relatively simple ways that the researcher can systematically organize this information.

4. Metadata Categories. A description of an archived resource is also known as its METADATA and is composed of many discrete pieces of information, or ELEMENTS. A metadata SCHEMA is a definition of the specific elements used, their precise meanings (including the ranges of meanings they might have), rules of use and relationships among them.[3] The OLAC Metadata Set is a proposed standard for a basic level of resource description to be used by data providers (archival repositories) in the Open Language Archives Community (Simons and Bird 2001). It is still a draft under development and not yet a stable recommendation, but it is a very good starting point for this discussion.

The OLAC Metadata Set (draft of 25 April 2001) contains 24 elements. Of the 24 in the OLAC set, five pertain specifically to software resource description and are not treated here. The prose descriptions here are intended to aid general understanding of their use and significance.

- **Contributor**: the name of an individual or organization which has contributed to the resource but is not primarily responsible for its creation. Information about a person or service which has performed physical conservation work, reformatting, or other work to make the resource more useful or rescue it from digital obsolescence might also be noted here. If possible, further specify the role of this entity, e.g. 'sponsor' or 'service bureau for reformatting'.

- **Coverage**: the spatial location or temporal period to which the resource pertains. If location can be reasonably predicted from the language identification, it is not necessary to specify it here, though noting the country provides a check on the correct assignment of language code if the name is ambiguous.

- **Creator**: the name of an individual, collective group, or organization which is primarily responsible for creating or compiling the intellectual content of the resource. For a sound recording, this would typically name the language consultant(s) or performer(s) recorded; for a transcription, this would name the transcriber. If possible, further

[3] Those who wish to delve deeper into this general topic will find a good introduction to metadata systems in McKemmish, Cunningham, and Parer 1999.

specify the role of this entity through a term such as 'performer', 'transcriber', or 'compiler'.

- **Date**: a date associated with an event in the history of the resource: date of creation, date of format conversion, etc. Where possible, specify year, month, and day, but if this is not possible, give an approximate date by year. More than one date (and event) might be needed in order to trace the development of the resource in a way that is useful to another researcher.

- **Description**: a prose description of the contents of the resource, such as an abstract, table of contents, or note about physical characteristics. Special circumstances surrounding the work can be given here, if the information does not fit in another element (such as format or type), or requires more explanation than can be accommodated in that element. For a collection of texts, this might give the number of texts or recordings, a generic physical description, and a list.

- **Format**: the physical medium or digital manifestation of the resource. This might be 'paper manuscript', 'PDF file', 'SFM file', or 'cassette recording' and can include the size or duration of the resource in pages, bytes, number of entries, hours/minutes/seconds, etc.

- **Format.encoding**: for a digital textual resource, it is critical to identify the character set used. A unique font and its encoding developed for a particular language will constitute a separate resource to be archived along with other resources.

- **Format.markup**: for a structured textual or multimedia resource such as an interlinearized text (perhaps done in Shoebox 4.0), documentation of the tagging used is essential for future users of the resource.

- **Identifier**: this is the means by which the repository will unambiguously identify this specific resource. It will be assigned by the repository, but the depositor may want to make a note of it for ease of future access. If you have developed your own identification system for your materials (and labeled them), the repository will also benefit from having this information, particularly if it helps to link related materials (recording with transcription, etc.)

- **Language and Subject.language**: These can perhaps best be distinguished as 'commentary language', the language in which analysis and/or description is given (or the language of the intended audience for the work) and 'language under study', a language that is the topic of

description or analysis. In order to avoid ambiguity, it is very helpful for the repository to have these (particularly Subject.language) specified by its Ethnologue code, or some other standard coded identifier for languages, such as ISO 639 (ISO 1998).

Though not yet incorporated into the OLAC Metadata set (or into the structure of its controlled vocabulary for language identification, which is the Ethnologue), it is also useful for the repository to have information regarding the linguistic family and stock to which this language belongs.

- **Publisher**: this pertains only to a resource that is already available to the public through some common distribution channel, such as a published dictionary. Publication (and distribution) of a work is one of the rights reserved to a work's creator, but sometimes transferred in whole or in part in the publication contract; information about the (previous) publication of a work is not only helpful to users wishing to obtain the resource, but also important to the repository in managing the resource and making it available in accordance with any intellectual property claims.

- **Relation and Source**: these two can be a bit difficult to sort out. The intent is to capture information about another resource that is related but independent (use of this resource is not strictly dependent on having the related resource). Use SOURCE where the intellectual content is derived from another resource but is now of a different TYPE and additional creative work was involved in making the new resource. For example, the SOURCE metadata for a transcription would list the sound recording on which the transcription is based. Use RELATION where giving information about a previous or succeeding version or different FORMAT of the content, such as a PDF file of a printed document.

- **Rights**: this is a statement regarding who has what rights or permissions to access, use, distribute, or make other works derived from this resource. Whatever is known about specific agreements made with the originators of the intellectual content of the resource (a language consultant, another researcher, a language community) should be made clear to the receiving repository.

- **Subject**: This contains keywords that describe the topical content of the resource. This might refer to a linguistic theory on which an analysis is based or a sociolinguistic or anthropological concept dealt with in the

work. This can also be used for a subject term or phrase for the topic of a discourse, such as 'canoe building' or 'healing ceremony'.

- **Title**: The formal name by which the resource is known and would be cited. For a title given in the language under study (Subject.language), it is also helpful to supply a gloss of the title.

- **Type**: This identifies the broad category to which this resource belongs, such as: Collection, Dataset, Graphic Image (such as for a photograph or JPEG file), Software, Signal (Sound or Video), or Text. For 'Collection' it is helpful to specify the type(s) of the items contained in the collection. A scanned image of a page of text is identified as Type: Text, since it is intended to be interpreted by the user as text rather than as a representational picture. In most cases TYPE can be deduced from the FORMAT information, but in some cases that is still ambiguous, as in the example of a scanned page of text which can be of the same FORMAT as a scanned photograph.

- **Type.data**: This element identifies the nature or genre of the resource in specifically linguistic or ethnographic terms.[4] Broad categories considered here are 'Transcription', 'Annotation', and 'Description'. Subcategories can be given to identify the type of transcription and/or annotation, e.g. phonetic or practical orthography, or the genre of a written text of recorded speech event, e.g. personal narrative, conversation, sermon.

It may be difficult to decide whether a particular bit of information belongs in one element or another. Provide whatever information you can for each of these, associate it as best you can with the most logical element and try to be consistent in your choice of element across your descriptions of different resources. In case of confusion, try to give more explanation (rather than assuming one meaning over another), and let the repository worry about fitting the bits into the proper boxes in the best way.

5. Seems like a lot of work... Documenting language resources may seem like it involves a lot of tedious work. (Congratulations on reading this far in the article!) That perception is accurate, but there are some ways to ease the burden. The best way to simplify the work of documentation is to do it as

[4] This category is heavily influenced by the interest and focus of the researcher and of the repository. A good example of this is The Archive of the Indigenous Languages of Latin America. AILLA is developing a metadata set specifically designed for their collection of recordings, transcriptions, and translations—mostly of naturally occurring discourse. Their focus is primarily on ethnography of discourse, and so the metadata which their system uses reflects and supports this focus. (Michael, 2000)

you go, and in a consistent manner. Use the elements of the OLAC set as a guide to what you should document for each resource you collect or create in the course of your work. Keep in mind that not every element is meaningful for every resource, or even every type of resource, and that other types of information not mentioned here might be important. Use a Shoebox file, a simple database, a Word document or note cards—whatever is easiest for you to maintain—ready to copy and hand over to the archival repository along with the resources themselves. Since this sort of record also serves as an inventory, it is helpful to have it separate from the resources themselves, rather than (or in addition to) embedded within resource headers or introductions.

Another very helpful approach is to use a technique that archives have used since the beginning of archival work as a distinct discipline: organize individual works with common characteristics into collections and then describe the collections. For example: among your resources you may have a lot of individual recordings on cassettes that are similar types of materials and are transcribed in pretty much the same ways in a group of computer files. These may have free translations in other separate files. Rather than describe each recording, transcription, and translation individually, describe them as three collections of resources. A specific case might be like the following (elements are reordered here in a more logical progression, rather than alphabetically):

Title	Collection of *Language-X* sermon recordings
Creator	*Name(s) of speaker(s)* (with their permission); if the speaker does not wish to be identified, it is helpful to have something like 'a *Language-X* church elder'
Contributor	Name(s) of individual(s) or group(s) that recorded the speech events (probably yourself among them)
Date	Time period over which the recordings were made
Language	Name and Ethnologue code of Language-X
Subject.language	Leave blank, since the Language element is sufficient
Description	53 recordings: 1. *Title. Title gloss. Date. Speaker. Length* 2. *Title. Title gloss. Date. Speaker. Length* ...
Coverage	Country where Language-X is spoken
Format	Cassette

Format.encoding	Not relevant for non-digital resources
Format.markup	Not relevant for non-digital resources
Type	Collection: Sound recording
Type.data	Sermon; Religious oratory; Hortatory speech as appropriate
Subject	
Source	
Relation	
Publisher	
Rights	The church elder who originally delivered these sermons wishes them to be available to members of his language community and to scholars, but requests that permission from the *Y Church* leadership be obtained before publicly airing any sermon or publishing any transcripts.

Title	Transcriptions of *Language-X* sermon recordings
Creator	Your name or other primary transcriber(s)
Contributor	
Date	Dates of transcription
Language	Name and Ethnologue code of Language-X
Subject.language	
Description	53 transcriptions of recordings: 1. *Title. Title gloss. Date. Speaker. No. of pages. Reference to match with specific recording* 2. *Title. Title gloss. Date. Speaker. No. of pages. Reference to match with specific recording* ...
Coverage	Country where Language-X is spoken
Format	Standard Format
Format.encoding	Identifier for the character set used
Format.markup	Description of sf markers used, or reference to where they are documented.
Type	Collection: Text

Type.data	Transcription/practical orthography
Subject	
Source	Collection of *Language-X* sermon recordings submitted with transcriptions
Relation	
Publisher	
Rights	The church elder who originally delivered these sermons wishes them to be available to members of his language community and to scholars, but requests that permission from the *Y Church* leadership be obtained before publicly airing any sermon or publishing any transcripts.

Title	Free translations of *Language-X* sermon recordings
Creator	Your name or other primary translator
Contributor	
Date	Dates of translation
Language	Language of translations
Subject.language	Name and Ethnologue code of Language-X
Description	48 translations of recordings/transcripts: 1. *Translated Title. Original title. Date. Speaker. No. of pages. Reference to match with recording* 2. *Translated Title. Original title. Date. Speaker. No. of pages. Reference to match with recording* ...
Coverage	Country where Language-X is spoken
Format	Rich Text Format
Format.encoding	ASCII
Format.markup	
Type	Collection: Text

Type.data	Sentence level free translation
Subject	
Source	Collection of *Language-X* sermon recordings submitted with translations (48 of 53 recordings were translated)
Relation	
Publisher	
Rights	The church elder who originally delivered these sermons wishes them to be available to members of his language community and to scholars, but requests that permission from the *Y Church* leadership be obtained before publicly airing any sermon or publishing any transcripts.

It is tempting for the researcher to conclude that much of the descriptive information could be 'figured out' by the archival repository once the materials are in their custody. This may be true of some elements of the description, but most of the metadata really cannot be deduced easily (or even at all) from the resources themselves, without risk of error and a lot of time spent eliminating other possibilities.

6. Archiving—a partnership: Archiving is essentially a partnership: the depositor and the repository working together to organize, describe, and preserve valuable resources in anticipation that someone will want to use them again in the future. In linguistic and anthropological research these resources are the products of years of labor and a significant relationship between a particular community and the researcher. Thus, a very important third—though sometimes silenced—member of the partnership is the language community and the individuals who have worked with the researcher. Often the repository must rely on the researcher to obtain and convey a record of their interests and desires relating to these products of both individual creativity and collective linguistic and cultural heritage. In its turn, the repository does its best to represent and enforce these interests and desires in the terms of use imposed on future users of the resources. Potential future users—members of the language communities themselves, other scholars, educators, government officials—will be looking for materials like these. The archival repository will have plenty of work to do to ensure that these resources remain viable and accessible. Getting off to the right start on this task will depend to a large extent on the descriptive information supplied by the people who have created them.

REFERENCES

Alvestrand, Harald. 2001. Request for comments: 3066: Tags for the identification of languages. http://www.ietf.org/rfc/rfc3066.txt

Bearman, David. 1999. Reality and Chimeras in the preservation of electronic records. D-Lib Magazine 5(4) http://www.dlib.org/dlib/april99/bearman/04bearman.html

ISO. 1998. Codes for the representation of Names of Languages-Part 2: Alpha-3 Code. http://lcweb.loc.gov/standards/iso639-2/langhome.html

McKemmish, Sue, Adrian Cunningham and Dagmar Parer. 1999. Metadata mania: Use of metadata for electronic recordkeeping and online resource discovery. In Place, interface and cyberspace archives at the edge, Proceedings of the 1998 Conference of the Australian Society of Archivists, Fremantle 6-8 August 1998. Canberra: Australian Society of Archivists 1999. pp, 129-144. Available also at http://www.sims.monash.edu.au/rcrg/publications/recordkeepingmetadata/sm01.html

Michael, Lev. 2000. Creating discourse data metadata for the AILLA Project: Lessons learned and needs felt. In Papers from the Workshop on Web-Based Language Documentation and Description, Steven Bird and Gary Simons, editors. http://morph.ldc.upenn.edu/exploration/expl2000/papers/michael/michael.pdf

Rothenberg, Jeff. 1999. Avoiding technological quicksand: Finding a viable technical foundation for digital preservation. A report to the Council on Library and Information Resources. Washington DC: CLIR, 35p. Available also at: http://www.clir.org/pubs/reports/rothenberg/contents.html

Simons, Gary and Steven Bird, editors. 2001. OLAC Metadata Set. (Draft standard) http://www.language-archives.org/OLAC/olacms-20010425.html

Joan Spanne, 7500 W. Camp Wisdom Road, Dallas, TX 75236. E-mail: Joan_Spanne@sil.org

■

BEST COPY AVAILABLE

The SIL Language and Culture Archive: An Interview with Joan Spanne

April 6, 2001

Why archive?

E: Joan, you have been entrusted with heading up SIL's Language and Culture Archive.[1] Why should we have such an archive?

J: There are two very basic motivations within SIL: one of them is to keep the product of our work safe, both what is complete and what is not yet complete, for our own purposes, that is, for the continuation of the work. The other is that we want to make material available for those outside of SIL, to satisfy the academic service part of our mission. An archival repository needs to have a specific mission statement, one that supports the mission of the organization. A lot of knowledge about the repository's context goes into decisions on what resources to collect, how to keep them and who may use them. I discuss that in an Archiving Guidelines document recently sent out to SIL entities.

What are Guidelines?

E. You mentioned 'Guidelines'. Are the Guidelines readily available to all the members of SIL?

J: They are available to anyone within SIL and to those who are closely associated as partners, to read and comment on. They are not specifically for the OWL, the ordinary working linguist, who might be overwhelmed by them. But there are those owls who are interested in it and who could make very helpful comments, and so the Guidelines are not closed or confidential in any way; probably the easiest way to get a copy is to contact me.

E: So they can request a copy by emailing you at joan_spanne@sil.org?

J: I can send a copy as an attachment, in a Word document, or for those who are working on Macintoshes, an HTML file.

---

[1]The interviewer (E) is Eugene Loos.

Can a linguist get credit for 'archiving'?

My CV!

E: Taking 'archiving' to mean the two ways that you have mentioned in the Guidelines, as 'storage for retrieval just in case something happens' versus 'storage for retrieval by those who want it', I would guess that most people probably would be glad to have the security side attended to by someone else. The part about 'making information available' really takes the burden off of them to 'get quotable'—make their work more broadly known. How can we go about seeing that an author really gets credit for what he/she has done? Someone might plagiarize what we have taken years to put together!

J: Until 1978, protection under the law for the original linguistic or artistic expression—meaning the copyright—was tied to publication, or at least to the registration of a work with the proper national authority. From 1978 onwards, the protection of copyright became inherent in the act of creation, so that unpublished or unregistered works are also protected. In other words, it is not only unethical, it is illegal for a researcher to publish as his own the work of another researcher. The Archives can help establish original authorship through the records it keeps of materials deposited and the measures it takes to protect those materials from unauthorized copying or theft.

E: So if someone uses something available in an archive, anyone can double-check what the user is doing to see whether he has given proper recognition for what he has used?

J: Yes.

E: That's a wonderful comfort, I think; an assurance for the field worker. I would think that teams who have been on the field 5, 10, 15, or more years might now have an accumulation that they would find it a fearsome task to catalogue.

Who is going to do the depositing?

It took me 45 years!

E: Just think of all the word lists and surveys and so on. Isn't that just too much to ask of anyone?

J: There are two answers to that question and they're tied together. It is a lot of work. It must be

recognized by the branch that preparing materials to be archived is important work, it's not 'if you have time'. The entity that is responsible for overseeing the work should make archiving an expected part of a field linguist's assignment.

Isn't description an onerous task?

E: OK, well, the Peru Branch has had an open archives policy so they microfiched everything to make it all available, but nothing was tagged. Perhaps the name of the text or something is there, but how would we go back and log in all that microfiched stuff—some of which is scarcely legible?

J: We have some of those materials in the L&C Archive because when the microfiche was made, copies of the fiche were sent to Dallas. And you're right; some of those are not very legible. But there is internal information that a knowledgeable person can get out of the text that helps to identify it.

E: That implies that someone would have to peek at it, to see what it is.

Where is the data?

J: Yes. In just about every case the researcher is going to need to have the material in hand to decide whether it fills his or her need. There is perhaps more information available concerning those microfiched materials than you are aware of. There was some basic indexing done here in Dallas, identifying as much as possible: the linguist involved, the language, the date, and the type of material that is there. Very basic, of course, but it tells more than that it is just a piece of paper or a set of microfiche, but in most respects it is going to be the responsibility of someone wanting to use the resource to do a fair bit of preparation to get the materials to the place where they can be used.

Archive what? Old stuff?

E: Well, let me ask about pages of data made long before computers were available. Maybe it is on old, yellowed paper, written in pencil; what can we do with that?

J: There are three basic approaches for that older material: one is to save the piece of paper. I'm actually getting to be a proponent of archiving paper. A second alternative is to film materials, as has been done by some branches. Peru is not alone in that. The third possibility would be to scan material. Both filming and scanning require equipment and some degree of knowledge

in order to do a good job. We've kind of grown wide-eyed at the possibilities for distribution, making material available on the Internet, because scanned material is so easy to send around. The difficulty that I see with scanning is that, once you've performed the scanning work, you must continue to maintain the image files, along with a lot of structural information about how the files are related to each other. This is in addition to providing the information needed in order to find the resource in the first place, which is needed for a resource in any medium. There are many different bits of information, and keeping those bits together in an ordered manner is a lot of work, as you already know from your experience in scanning jobs that you have done. The same is true with microfiche, except that microfiche connects the material together in a physical medium that is much less subject to technological obsolescence. They are at least readable if you have a light source and a magnifying glass. The information that needs to be collected simply to preserve the work in a hardcopy form, whether still on paper or on a film, is not nearly as intensive as the work in uniting and preserving digital materials.

E: What about early work? Everybody has first transcriptions, made perhaps before they settled into the phonemic patterns. What would you recommend concerning those early word lists and texts?

J: That's a tough question, an evaluative question that I as archivist would have a very hard time deciding on. That's where I would find experienced linguists to ask for an opinion of the value of the material. I would say that if it is in a language in which there is a large body of materials that have been made available to publication or archival collection, then probably early work that might not be as reliable, either in transcription accuracy, or because of orthographic changes that have taken place, would not have a great value. We would have to look at the value of it in comparison to alternative sources that there are for data analysis.

E: I have had the interesting experience that sometimes in looking at tentative word lists of languages for which there is nothing much available, if you can get a comparative set of early transcriptions by different people attempting to record the same thing, sometimes you can discern phonetic features that otherwise might not have been captured. Those are cases where early, rough stuff is quite useful.

J: One thing that immediately jumped out to me in your statement was THERE IS NOT MUCH AVAILABLE, and so the fact that it is early work may make it less reliable in some situations, but as you point out, it depends on your purpose, your use for it. The fact that there isn't a whole lot available increases its value. For a language that has very few or no remaining native speakers that was studied perhaps 20, 30, or 50 years ago, and therefore we'll have no one collecting material anymore in that language, what we have is what we have. Then, yes, those types of materials could be very valuable, but again, I would look to a knowledgeable linguist who knows what has been produced in that particular area. If he/she says, 'Yes, this is rare stuff', we need to keep it.

It will be safe!

Archive in what form?

J: Back to the question of medium, I'm more and more a proponent of keeping paper copies for some types of work. While working with Central America Branch's material we had collections in paper copies and the digital files that were somehow related to the paper copies; either the hardcopy was typewritten copy that was later transcribed onto the computer, or in other instances we had the output produced ten years ago from computer files. We had to do a lot of going back and forth between the source and something produced from that source. Producing the files from some of those older computer files took a fair bit of effort, but it was not impossible. There are some problems with those computer files, characters, and fonts to be rendered; in some cases we could figure out what they were originally, to get a one-to-one correspondence, in other cases we're not really sure. So a lot of that depends on the documentation that goes with those files.

E: Like associated cc tables and fonts or printer drivers?

J: Sometimes cc tables; in other cases a description of the process that was gone through in converting the materials because sometimes the files that we have now are actually conversions of much older materials, a very mixed bag. But with typescript originals we can pretty well make out what the author intended. When the interested party wants to get at the intellectual content of the materials, to read it and understand it, the paper is as good as or better than the computer files. When researchers want to put the material through some kind of further analytical process, usually done by computer these days, they want the computer source files. So they're going to have to put the material that we have in print into a digitized form in order to apply

the desired analytical process to them. But a journal article, a conference paper, a report on findings, descriptions of analytical work done, are primarily documents for which the intellectual content is what the researcher would be interested in and so for those I think it is most appropriate to have them in hard copy.

E: If you keep things in hard copy that comes from all the branches, you've got a storage problem!

J: Yes, we do have a bit of a storage problem, but we are now talking about unpublished materials that are of such a nature that we don't have the motivation to do the extra work of converting to computer files. Our Scripture materials are already archived in the printed form and also in computer form where such exists. We are committed to maintaining in the archives here any paper copy or microfilm that the branches have produced and we already have. And so when we eliminate that which has already been published, we've narrowed down the scope of the task for material on paper. There are types of material that we have strong motivation to keep in digital format, or to convert to digital format. Those would be language texts and data files, lexical files, files of field notes, recordings. How we deal with those is a whole separate question because we do have a high motivation to preserve them in digital format for future use. Our shoeboxes of 3X5 notes are far less usable, less accessible to the researcher than a lexical file in a computer database.

E: I still want to pursue the idea of preserving hard copy. If you collect hard copy, how do you attend to the need of the researcher who must access it?

J: In a lot of cases the researcher will be expected to come to Dallas. It is normal that a researcher must go to the place where he/she can obtain access and view the materials rather than, say, just requesting wholesale sets of materials. In other cases we will make photocopies available, because when we reach the stage of a researcher actually requesting specific materials, the motivation for doing something with those materials jumps up considerably. One of the reasons why I'm interested in preserving paper copy is that while there is definitely a space cost, there is a considerably lower labor cost, and labor is expensive, even in our organization. We are not used to counting the cost of the labor that processing requires. When we look at the cost and compare it to an unexpressed, unknown interest in the future, we have a hard time committing resources to the labor of putting it in a more usable format. When a request comes in for a body of materials for a language, then the cost benefit for dealing with that particular language goes up, and so in many

instances it can become a partnership between the archives and the researcher to make the material available in another form, and to work with that researcher who then makes an investment in the material and benefits from having access to it, which also returns some value for the benefit of future researchers.

E: By making necessary copies, cataloguing, and so on?

J: Yes. Even by hiring someone who actually keyboards material, then committing those computer materials to the archives, which can make them available. That way the archives actually participate in the production of the reformatted materials. The work of maintaining materials is greater than just receiving materials from an outside source. The archive needs to be in control of the format and the documentation from the structural and administrative perspective.

E: I can envision that some researchers would be interested in having all the materials related to a particular language family. Those materials might not be currently accessible in any bundled set. But if he/she were to undertake the bundling of them and make that bundling be his/her contribution to the archive, it would be a very useful contribution—a way of utilizing efforts.

What about updateable stuff?

E: Would you archive something that possibly is going to be updated in a couple of years?

J: A couple of years is on the margin of when I would say to archive it, because, the Lord willing, you'll be able to complete the work. Yet, who really knows what is going to happen? It kind of depends on the project. If it will be five years before the next update, certainly, document it and put it in the archival collection. If it is something very much in process, or if you expect to get back to it in eight months, it's really not in a form that you would want to make available to other people. In those cases I would say that is the kind of material that most assuredly you'd want to back up, not archive.

E: If a person envisions that the material is rather static now but wishes to update later, will he have assurance that he can replace it with an updated version?

J: Yes, we will need to be able to replace superseded material. In some instances the entities themselves might express the desire that the material not be updated, that subsequent versions just be added, to keep track of

versions made. That's a special case. Usually it will be appropriate to remove the earlier materials because they're not going to be of value for their linguistic content after the updated version is available.

J: The tension between two things—work in progress versus work at a hiatus—is one of the reasons why a good backup process complements the archiving process, because you find a spectrum of materials. At one stage work clearly belongs in a backup, not intended for anyone else, and then it moves further along to the stage where it is not quite ready for publication but is completed to an archiving level even if not ready for publication. It is a judgment call.

How much time might be needed for archiving

E: How much time to you think it might take for the field worker to archive an accumulation of data?

J: That's a loaded question. A lot depends on the documentation that is already there, made right up front when the materials were created. If you have to rack your brain, asking 'What is this?' it will take more time than if you can find a clue and say 'Ah! Look at this! I created a header that told me who this is, when I did it, the context of (if it was a sound recording and a transcription), who assisted me in the transcription,' and 'Oh, that helps me because that person did something in this way, whereas another of the people did transcriptions in another way.' That's one reason why it's important to think through the whole process from the beginning, which is difficult to do. You may be fully occupied doing your particular task and don't want to be bothered by what use someone is going to make of it later.

Can we compress before archiving?

E: Some things, sounds and images in particular, are very voluminous on computer. So how can we expect to accept such voluminous stuff? We would soon run out of space in which to store it!

J: Archives running out of digital space really isn't that much of a problem. Bytes for storage are just not that expensive anymore. They can be expensive to transmit, and so we have a desire for smaller, kinder, more compact processes of information when we want to transmit it over the Internet. Think of both an archival information copy and a distribution copy. The distribution issue may be solved by using a compressed format such as PDF or DjVu™ or some other compression method. A couple of weeks ago colleagues asked about putting materials on a CD and using PDF because of a request that they'd received from a university. I said 'Yes, by all means, do that.' Especially if they're going to support you in doing the job. He will be making a product that will be useful to himself and more accessible to a wider audience. It fulfills a need and a desire of an institution with which one wants to develop a relationship. But don't consider that to be archiving. That is a product, something that you produce from materials that you have archived in, shall we say, more secure means, in terms of preservation strategies.

Eight Gigabytes on one diskette!

E: You mentioned PDF and DjVu.[2] Both are means of bundling and compressing files. My experience with PDF is that you can't get out of a PDF file the same quality of picture as when you print out an original BMP or TIFF file, especially if the PDF file is calibrated for the computer screen rather than a high-resolution printer.

J: That is true. And that is one of several reasons why PDF isn't considered to be an archival format. It is a distribution format.

---

[2]J: PDF is a file in Portable Document Format, produced through a software product called Acrobat, from Adobe. Acrobat is used on a wide variety of text and files to produce something that packages together paged information that has very much the look and feel of a book or document on paper. It can be read on a variety of platforms using a software reader, so a PDF document is acceptable to Macintosh, Windows based systems, or UNIX based systems. In order to be able to open the document and read its contents all you need is an Adobe Acrobat reader. DjVu is another technology that has a freely available reader for Windows and Macintosh. DjVu is based on a digital imaging format designed specifically for capturing scanned images and publishing them on the World Wide Web. AT&T Research Labs originally developed the technology, which is capable of very high compression ratios. DjVu is a registered trademark of LizardTech, Inc. (http://www.lizardtech.com/), which offers a suite of products for creating DjVu documents. A freely available reader can be downloaded at http://www.lizardtech.com/cgi-bin/products/desc.pl?tsb-25720.

E: That brings up the question of whether a field linguist should use something like ZIP to compress and bundle their files together first.

J: The reliability of compression technology like ZIP is of sufficient level that I would accept into the archive materials that are zipped for transmission so as to make them as small as possible to get them to the archive. But for storing them in the archive, we won't keep them in zipped format. We will unpack them and store in the native format, after checking to see that they are intact.

E: I suppose that most of us have made less-than-desirable recordings because a recording machine added noise or for other reasons the recording turned out to be less faithful than one might have desired. The recording can be cleaned up with something like COOLEDIT.[3] Would you recommend people clean up their files before submitting them?

J: I would recommend that they send us the original. Or that they work with a reliable archive in doing any preservation work, including digitization of those files, partly because the more direct control that the archive actually has over conversion from a fragile medium such as tape into another format, the more they are going to know about the history of that material, and the better they are going to be able to preserve it and to document the action taken to preserve it. The documentation of those actions can be done, of course, by the depositors, and so I would say, 'Participate with the archive.' If it is something that you want to be involved in, great! The Language and Culture Archives would like to have the involvement of the individual researchers, the linguists, in the preservation of the material. I would like to be able to say that the Language and Culture Archives is in a position to act today on that preservation work. We're not, but that is a significant project that I'm working on right now.

Access to the archive

E: That's encouraging. Who will have access to archived material?

J: That depends on which specific material or specific collection is involved. While a language program is in existence, and also while there is an entity supervising the work in the area, country, or region where that is going on, then it is the entity or supervisory staff that establishes any controls over who has access to the materials. Even if deposited materials originate in

---

[3]CoolEdit is a product of Syntrillium Software Corporation.

Language Project A in Country C they are still controlled by the parameters that are set by that entity or staff; there may be issues of sensitivity. If there is a question of a researcher intending publication within a reasonable time frame (say, 5 years), then those materials will be protected according to parameters set by the entity from which the material originated.

Authorization to distribute

E: I think that a significant use of the archive will be for comparative work in linguistics and anthropology, and so getting together all the available material, at least the unpublished materials, is a tremendous service. I have received a number of personal requests for materials with information on languages of the Panoan language family, the family to which Capanahua belongs. As years went by on the field, I latched on to Panoan word lists whenever possible. Later when I received requests for copies of those lists, I had to ask myself if those particular materials were authorized for distribution. So I had to refer back to the entities they originated in and get clearance to give them out. If a person gathers materials that are not necessarily his own, just gathering them because of interest, would we archive it?

J: Well, that's a very common situation, but when it comes to redistributing them, yes, you are right. There is a range of interests involved in those materials and that's one of the gray areas to work out, in terms of documenting how soon they may be used, who has right over the material, and how those should be acknowledged—both in terms of the ethics of acknowledging another's work, and also the legal aspects of the right of reproduction, distribution, and the right to create derivative works—all of that sort of thing.

Published materials

E: A field linguist's material might include things already published. Will we archive published materials? And the source files of those publications?

J: Materials published by SIL entities, including SIL International, have been archived for years, originally by the SIL Bibliographer, and now in the Language and Culture Archives. Academic Affairs is working to clarify the relationships, rights, and responsibilities of SIL workers, SIL entities, International Administration, and our partner organizations regarding academic publishing and republication of works. Works by SIL authors, which have been published outside SIL, are also archived by the L&C

Archives; we can only provide a copy of such a work in accordance with 'Fair Use' provisions of copyright law.

E: That should be assuring for depositors. Let's say that a book of native texts of a particular language has been published, and somebody wants a digitized copy of that, but the book is still available. What will be our policy; will we say he may not have the digitized copy because the published copy or book is available?

J: Access to digitized copy of materials published by SIL is a marketing issue that Academic Affairs will address. Copyright rules determine how much access there may be to materials published outside SIL.

Alternative archives

E: You know, there are other similar archives under the auspices of different institutions; would we say a person should archive in some other place as well as SIL, or only in one institution?

J: That depends partly on the capacity of the institutions to provide an archival environment that is secure for the long term and dedicated to making materials available. If that archive is going to be able to preserve the material and make it available, it's appropriate for an SIL entity to develop a relationship with that archive. Such is the case in Australia, where the AAIB has more than a ten-year history of depositing material with a particular repository. They have committed to depositing material there; consequently we don't have much of those materials here. They have been maintaining materials in a branch collection but that branch collection isn't necessarily going to be transferred wholesale to another institution because they've had this progressive program in Australia. In some other countries where we've worked we have deposited materials with various universities and agencies, which is something we ought to do. However, there is no assurance that the materials will continue to be available. So when the level of confidence for long term preservation is not good, the entity should make efforts to deposit their collection with an established repository, whether its the L&C Archive in Dallas or another archive, where there is that level of confidence. In a situation where multiple copies have been deposited in various repositories, it should be clear among all of the repositories as to what can be done with those materials, what level of access there is, what rights over those materials exist, and who must be contacted to give whatever further permissions might be needed.

E: In the case of Australia, for example, will our archive have a record of where and what materials in Australia have been archived there, and be able to point researchers to that location?

J: Yes, a great question. When materials are deposited with another archive and that is to be the archival collection for long term, then a record of those materials must also be given to us in Dallas, so that as we receive requests we can direct researchers appropriately. Making that inventory is a part of the process of formally depositing. In a lot of places the language projects work with some degree of autonomy, and it may be the personnel in the language project that develop the relationship with an archive. That is great, we need that to happen, but the agreement is formally between that SIL entity and that other organization. Part of that agreement is a thorough listing of materials that are deposited. I don't mean in disgusting detail; I mean a listing of the collection contents that would be adequate to point researchers in the right direction, as well as be an adequate record placed in Dallas of what we actually deposited there.

E: So our archives will not only 'hold' stuff but also it will have 'pointers' to stuff?

J: Right.

Joan Spanne, 7500 W. Camp Wisdom Road, Dallas, TX 75236. E-mail: Joan_Spanne@sil.org

■

BEST COPY AVAILABLE

Several pages of the following review article, *The Amazonian Languages* edited by R.M.W. Dixon and Alexandra Y. Aikhenvald (reviewed by Terry Malone), were inadvertently omitted from the last issue of *Notes on Linguistics*. We reprint the entire article here. Please use this issue in referencing this article. We apologize for the inconvenience.

—Editor

REVIEW ARTICLE

The Amazonian Languages. Edited by R.M.W. DIXON and ALEXANDRA Y. AIKHENVALD. 1999. Cambridge & New York: Cambridge University Press. Pp. xxviii, 430. Cloth $69.95.

Reviewed by TERRY MALONE
SIL—Colombia

0. Introduction. Until recently modern linguistic theory, as practiced in the U.S. since the appearance of Chomsky's well-known 1957 and 1965 books, has taken little or no note of Amazonian languages. Two distinguished linguistic scholars have commented on this situation: 'When we began working together ... in the late 1970s ... there were scattered individuals in a number of countries in South America, mostly members of Christian missions, who were studying individual Amazonian languages, but general linguistics was being practiced almost entirely without reference to even the existence of Amazonian languages' (Derbyshire and Pullum 1998:3). This was in spite of sporadic attempts to challenge theory on the basis of data from Amazonian languages (for just a few examples see Pike and Kindberg 1956, Pike 1964, and David Payne 1974). Most, but not all, of this work tended to appear in out of the way publications, theses, dissertations, and specialist journals ignored by the theoretical mainstream.[1]

To be sure, some work in generative phonology had already begun to take notice of languages in the Amazonian area. The work was mostly related to developments in metrical theory (several works of Bruce Hayes), or in specific languages (work such as Everett and Everett 1984a,b). Earlier work

[1]The mainstream has equally overlooked the work of non-missionary linguists. One good example is Kaye (1970), who in spite of a short time in the field managed to produce a good partial description of Desano (Tucanoan) morphology. (A more complete Desano grammar written by a missionary linguist with 35 years of experience in the region is now in press.)

in autosegmental theory, chiefly with respect to nasal prosody within the more general context of harmony systems, had taken a passing glance at languages of the region (see, for instance, Safir 1982, van der Hulst and Smith 1982 and Hyman 1982). These developments inspired theoretical linguists to dust off works such as Bendor-Samuel (1960, 1966), and Kaye (1971), or scrounge around for work such as David Payne (1974) and Smith and Smith (1971),[2] but it was developments in typological theory, in part as a reaction to some of the failures of transformational-generative theory in the mid 1970's, which awoke current linguistic consciousness.

A crucial step in the awakening process transpired when 'in 1976, a professor in London was expounding on why no object-initial languages existed in the world. A student in the class hesitantly raised his hand and said, "Excuse me, but I *speak* an object-initial language." The professor was Geoffrey Pullum, and the student was SIL member Desmond Derbyshire' (Cahill 1999:1). This incident was the beginning of a series of publications (Derbyshire 1979, 1981,1985) which in interaction with the burgeoning field of typological studies has done more than any other previous development to bring the ignored and unknown languages of the Amazon languages to the lagging attention of theoretical linguists.

A simultaneous, even more significant step was the development of a fruitful editorial collaboration between the professor and his former student, beginning with the work Derbyshire and Pullum (1981) and eventually resulting in Derbyshire and Pullum (1986, 1990, 1991, 1998). At roughly the same time the two got underway producing these volumes, developments recounted in the introduction of Doris Payne (1990) transpired: thanks to some professors and graduate students at the University of Oregon who had worked in the Amazon region the complex classifier systems in some of the languages of this region were brought to the attention of theoreticians, first in Doris Payne (1986) (in the now classic Craig 1986), and then in two articles in Doris Payne (1990).

The next and still current surge of interest in Amazonian languages had its beginning with a renewed interest in the phenomenon of ergativity, most notably brought to the foreground of current theoretical consciousness in the papers of Comrie (1978), Dixon (1979) and the book by Plank (1979). Amazonian languages do not seem to enter into these earlier discussions; the year 1985 is the earliest date listed in Dixon (1994) for references discussing ergativity in Amazonian languages (in spite of earlier buried, provisional work such as Derbyshire 1983). Derbyshire and Pullum (1986), subsequent

[2]Early work such as Gomez (1980) has yet to come to the attention of theorists.

work by Derbyshire, and the involvement of the typological school at the University of Oregon (notably by the then Ph.D. student Spike Gildea) share major responsibility in bringing Amazon ergative systems to the attention of the wider linguistic world.

Some of these efforts came to the attention of the senior editor of the book under review, who was in part responsible for the simultaneous upsurge of interest in ergativity. He had, in his own words:

> ... devoted several decades to searching for substantive linguistic universals. In case after case, just as he thought he had achieved some significant typological statement, a counter-example popped up; and this was invariably from a language of Amazonia. He decided that the most sensible course of action was to learn Spanish and Portuguese and then go to South America ... In this way he achieved a degree of insight into the most complex linguistic area in the world today' (3).

He also ran into the coeditor, who had in the late 1980's begun researching obscure Brazilian Amazonian languages. One happy result is the book under review; another is the emergence of a second professional team who, like Derbyshire and Pullum, enjoy similar prestige in linguistic circles, and who can hopefully help to push forward what the productive Derbyshire and Pullum team has begun. 'The Amazonian Languages' is Dixon and Aikhenvald's major step in that direction.

A casual reader who is not a specialist in the study of South American languages and who is not aware of these theoretical developments might be tempted to pass by this book. That would be a crying shame: in spite of the title, the book is relevant to linguistic investigators in all fields, because it is hard to stumble across an Amazonian language that does not have something to offer to theoretical issues currently in vogue. Not only does the book provide an overview of the major language families of the Amazon Basin, but it also describes minor language families and isolates. In addition, it gives a good basic (if somewhat brief) introduction to the area and study of Amazonian languages. Two of the chapters describe synchronic language contact situations. Almost all of the chapters provide historical background for the study of each family or group of isolates, and all of them reveal areas relevant to current issues in phonology, morphology, syntax, semantics, comparative and diachronic linguistics, and grammaticalization.

In this review article I briefly recapitulate the contents of each chapter, at the same time emphasizing characteristics unique to or unusual in the languages being described, and relevance to current issues in the field. After the content summary I comment on some more general issues which the book raises, and then evaluate the book as a whole.

1. Contents. The editors first discuss conventions used in the book, including spelling, naming of language families, definition of 'language', grammatical terminology, following what Dixon calls 'basic linguistic theory', i.e. 'the accumulated tradition of linguistic description that has evolved over the last 2,000 years' (xxvi). This is a welcome section, too often missing in linguistic books.

In the first chapter they provide a good basic introduction to the book, discussing topics such as the purpose of the book, the situation of scholarship in the region, 'cultural background' of the region, 'linguistic diffusion' (areal linguistic features), proposed and likely genetic relationships, 'the punctuated equilibrium model' (see Dixon 1997), and the organization of the book.[3] The section on current linguistic scholarship and 'cultural background' are quite adequate for a linguistic anthology (that is what this book is), but anyone who desires or needs a more in-depth treatment of the cultural/historical situation of Amazonian languages should proceed to the excellent in-depth introductions in Derbyshire and Pullum (1986, 1990, 1991, 1998), where such information is more appropriate.

Although the author of each chapter has been given considerable freedom to describe language families according to the character of the respective linguistic systems, each chapter follows a similar format. Most begin with a brief summary of the history of studies in the family under consideration (including discussions of available classifications and comparative reconstructions), all provide maps of the location of language groups with accompanying population statistics, and all summarize the phonological traits typical across the family's languages. Each author then launches into a description of the language family's grammar. It is here that approaches differ significantly; nevertheless, morphology figures strongly in the majority of families, though it does not always appear under that title. Syntax receives attention where it is appropriate, and some authors even manage to squeeze in limited information on discourse grammar (text linguistics). Unfortunately, the authors had to observe rigid constraints on length for reasons outside of their and the editors' control; in spite of the incomplete descriptive coverage of Amazonian languages at present, even now a full book could be written on any of the seven families represented in the first eight chapters.

The first four chapters after the introduction cover the three largest language families in Amazonia, both in number of languages and of speakers: 'Carib',

---

[3] Readers who are familiar with the writings of both editors will detect the fingerprints of the senior editor all over this introduction.

by Desmond Derbyshire (23-64); 'The Arawak language family', by Alexandra Aikhenvald (65-106); 'Tupí', by Aryon Rodrigues (107-124); and 'Tupí-Guaraní', by Cheryl Jensen (125-164). Curiously, Arawak, the largest language family, comes second in the book, but otherwise the order reflects the relative importance of these families both in the region and in current studies.

Derbyshire's chapter on the Carib family is what one would expect from a scholar of his calibre; his mastery of the details in this summary reflects over 40 years of exposure to Amazonian languages. In his review of available classifications he notes that 'Carib comparative and historical studies lag far behind those of the other two large Amazonian language groups' (25). He does not dwell long on phonology: in Carib languages: the more interesting area is the interaction of phonological systems with the morphology, and it is here that phonologists will find plenty of fuel for theoretical fires, especially when considering the highly inflected verb systems.[4]

On the current linguistic scene Carib languages are most noted for their unique ergative systems. Because 'ergativity ... to a greater or lesser degree governs the case marking, person marking, derivational processes and constituent order patterning' (55), Derbyshire discusses it under several headings, including 'inflectional morphology' (31-37), 'main clause structure' (55), 'subordinate clause structures' (56-57), and a special section 'ergativity' (60-61). There are five 'dominantly ergative languages' (61), and a variety of split systems; the now well-known OVS (actually OVA) order occurs in ergative languages of this family. As in the Mayan language family (also known for ergativity), there is a close relationship between possessive nominal prefixes and verbal person affixes.

Derbyshire discusses at some length a current issue in the study of Cariban ergative systems: the two most distinguished scholars working in this language family differ on the origin and direction of diachronic change with respect to ergativity in the family. Derbyshire has argued elsewhere that earlier ergative systems are changing to accusative systems (Derbyshire 1991, also in the bibliography of this article), whereas Gildea (1998) argues that accusative systems are changing to ergative systems. Derbyshire provides a fair treatment of both sides, noting that 'Gildea's research has been extensive and his diachronic approach is sound and persuasive' (60), but at the same time he concludes that the issue cannot be satisfactorily resolved until 'fuller descriptions become available of more Carib

[4] For a sampling, see Gildea (1995), also in Derbyshire's bibliography.

languages', which in turn will allow more extensive comparative studies and 'a more reliable internal classification of the Carib family' (61).[5]

Cariban languages offer other interesting characteristics for typologists, phonologists, and Chomskyan theoreticians: locative postpositions inflected for liquids, flat surfaces, open areas, and enclosed places (42-43); a brief mention of 'the particle word class', which Derbyshire insists are not clitics but instead phonological words (53); and a rich array of nominalizing suffixes functioning syntactically as complements or adjuncts (43-52).

In a little over a decade Aikhenvald has managed to conduct extensive field work in Arawakan languages (mostly in south-western Brazil), and has an impressive command of the available literature. She is the only author who singles out endangered (vs. nonendangered) languages in her list of languages for the family (67-71), though comments on endangerment and survival prospectives for the respective families can be found sprinkled throughout the book. In her discussion of available classifications she notes that 'the first truly scientific reconstruction of proto-Arawak phonology ... was published by David L. Payne (1991). However, his subgrouping of Arawakan languages, which is based on lexical retentions, rather than on innovations, remains open to discussion' (74). David Payne has done work on possible shared morphological innovations across South American languages (David Payne 1990), so that very likely he used retentions in an attempt to filter out the effects of pervasive areal diffusion. Aikhenvald's evaluation here, as well as her own work on areal diffusion in the Amazon (one sample can be found later in this book), raises the more general theoretical issue of how one can dependably classify languages and trace genetic relationships in areas where extensive diffusion has taken place.[6]

---

[5] Derbyshire says that 'my view of the direction of change in the Carib family has been reinforced by a more general factor: the rampant ergativity that is found in so many Amazonian language families' (61). Ergativity is actually more wide-spread in Central and South American languages: I mentioned the Mayan family above, but languages in the Chibchan family (geographically intermediate between Mayan and Amazonian territory in pre-Conquest times) also offer a variety of ergative features in a generally decreasing cline to the south (see Quesada 1999 for a summary), and the Chocoan languages in western Colombia also sport ergative traits (Harms 1994). As one moves down the Andes, beginning in southern Colombia one runs into the thoroughly accusative Quechua languages. In evaluating Derbyshire's and Gildea's positions, one should keep in mind that Cariban and Arawakan languages extended further to the north in pre-Conquest times.

[6] My unpublished classification of Tucanoan languages mentioned in Barnes' chapter in this book also uses shared retentions instead of innovations. I have put the manuscript aside, chiefly because the present pattern of spreading for innovations in the family is clearly due to language contact, which in turn suggests that I have reconstructed a former language contact situation, instead of a network of genetic relationships.

BEST COPY AVAILABLE

Aikhenvald's phonological summary (75-80) suggests that Arawakan languages have much to offer theoreticians in the area of phonology, including glottalization, aspiration and nasalization as word prosodies (79), and interactions of morphology with stress systems (for instance, in Asheninka 'monosyllabic verbal roots have an obligatory prefix or a suffix, to make them bimoraic', 80). In fact, some of the stress systems have much exercised phonologists: Hayes (1995:288-296) presents a detailed reanalysis of the analysis in J. Payne (1990), concluding that 'this does not exhaust the stress phenomena of Asheninca' (Hayes 1995:296), and mentions the language repeatedly throughout his book. Asheninca stress figures in several papers in the Rutgers Optimality Archive, and in other current literature; it has become a testing ground for current phonological theory. Other languages in this family appear to offer theoretical challenges in this area: for instance some, such as Achagua, assign stress according to grammatical word class.[7]

Arawakan languages also offer much of interest to typologists: inalienable and alienable possession (with 'cross-referencing prefixes', 82-83); unique, complex, multiple classifier systems in interaction with gender systems (some languages have three interacting classifier systems, and one, Palikur, has five) (83-84); and complex verb morphology, including cross-referencing prefixes and suffixes (two-thirds of the languages), split ergative systems (typical of most Arawakan languages with cross-referencing suffixes), abundant valency-changing derivations, complex tense-aspect, modality, directional, and aktionsart systems (82-94).

Rodrigues is a fitting author for the more general description of the Tupí language family; he has over 40 years of experience working in the Amazon, and is currently the leading Tupían comparativist. Excluding the Tupí-Guaraní branch Tupían languages have been poorly described, or not described at all; Rodrigues makes do with what is at hand and provides an overall introduction to the entire family. Outstanding characteristics of these languages include inalienable and alienable possession, strict transitivity-based verb classes, 'rich systems of demonstratives' (120), 'subordinate clauses ... achieved through nominalization' (121), and some ergative characteristics (chiefly 'pivots', 121).[8]

---

[7] I base this statement on Achagua data that I have seen and analyzed. A phonology is in rough draft, coauthored by a Colombian linguist and an expatriot colleague.

[8] See Dixon 1994 for discussion of 'pivots' with respect to ergative systems.

In her chapter Jensen, who has over 20 years of experience in the Amazón, and who is also a Tupían comparativist, presents a summary for Tupí-Guaraní, the best known and best described branch of the Tupían family. Her historical-geographical introduction repays careful study; of all the families described in this book, European colonists learned to speak and used a few of the more prevalent languages exclusively from this family in preference to their native tongues (see 125-133). She devotes a considerable amount of space to Tupí-Guaraní phonology (133-145), perhaps best known among phonologists for a variety of unique nasal prosody systems, some of them bidirectional (135). Other traits of interest to phonologists include vowel epenthesis or consonant loss across morpheme boundries (the latter interacts with metathesis) (136), and voicing of bilabial consonants at morpheme boundries (137). Word-final consonants can devoice, disappear, or become nasals (142-144).

Traits of interest to typologists include split ergative systems interacting with the person hierarchy (O has precedence over A). The system includes four complex set of person markers interacting with transitivity and local discourse topic while marking A (subject of an active transitive verb), O (object of a transitive verb), S_a (the subject of active intransitive verbs), and S_o (the subject of stative intransitive verbs), 'the genitive in nouns', and/or 'the object of postpositions' (146-148).

The next three chapters cover four less extensive, smaller language families which nevertheless are significant in Amazonian linguistics: 'Macro-Jê' by Aryon D. Rodrigues (165-206); 'Tucano' by Janet Barnes (207-226), and 'Pano' by Eugene Loos (227-249). Rodrigues emphasizes comparative studies in his chapter on Macro-Jê (165-166, 198-201), because the internal consistency and external relationships of this group have been much questioned by area specialists. Nevertheless, there are also goodies which will interest the non-specialist, such as nasal prosody systems (in most of the languages oral and nasal vowels contrast, 171-174). One language (Karajá) has men's and women's speech (176-178). Plural marking is parsimonious (non-existent in the Jê family), sometimes appearing on pronouns but usually not on nouns; in some languages plural (A or O) is marked within the verb stem (183-184). Some ergativity occurs; Rodrigues' summary suggests that more work is needed to determine its exact nature (193-195).

Barnes' chapter on Tucanoan languages reflects roughly 30 years of experience in the Vaupés of Colombia, and immediate access to dozens of linguists who have similar experience studying these languages. She discusses Tucanoan classification using an unpublished manuscript of mine, basically an update of an older classification (Waltz and Wheeler 1972, in

her bibliography). All Tucanoan classifications and reconstructions are clouded by hundreds of years of extensive internal diffusion between Eastern Tucanoan languages, raising the theoretical question of how one can classify languages within a family, and reconstruct a proto-language in such a situation. The sociolinguistic situation in the Vaupés is unique among all contact situations in the world; more details can be found in references in Barnes' bibliography and in Aikhenvald's chapter in this book on diffusion and language contact in the Içana-Vaupés area (see below).

Tucanoan languages have much to offer the typologist and phonologist: they sport the most complicated, highly inflected evidential systems in the world (at least among the world's languages that have been studied); most have rich classifier systems (suffixes on numbers, demonstratives, nominalized verbs, and some nouns) in interaction with a gender system; almost all have nasal prosody; and across the family a variegated array of poorly described pitch-accent or accent systems begs for comparative study. Nasal prosody in Tucanoan is difficult to describe without postulating 'three-way autosegments;'[9] in spite of some interesting proposals within Optimality Theory, it is questionable whether this theoretical quandary has been resolved. Curiously, the Tucanoan languages are all thoroughly nominative-accusative in a region where ergativity is dominant.

Loos' chapter reflects over 40 years of experience with Panoan languages. Panoan phonology is complex (230-234); it is perhaps most distinguished by 'nasal spread' and complex interactions of morphology with the metrical system, i.e. 'an odd-even syllable-timing characteristic common in Pano languages causes phonological modifications such as segment deletion, plosive nasal release, stress assignment and possibly vowel harmony' (232).

For typologists there are 'a variety of split [subject-marking] systems, the marking of the A and S being affected by focus' (236) within a system of 'transitivity concord' (some subordinate clauses, some adverbial verb suffixes, locative phrases locating A or S, and certain sequential clauses must be marked according to verb transitivity). There is a complex switch reference system, perhaps the most complex in the Amazon, and best exemplified by Sparing-Chávez 1998 (described for Amahuaca, not available to Loos at the time of writing). Ergative marking occurs, uniquely marked by a syllable final nasal /n/ which often disappears and leaves nasalization as its only clue; in turn, 'in some of the languages the

[9] Barnes lists a paper in her bibliography describing how these segments operate (Barnes 1996), both for pitch and nasal prosody in Tuyuca, but she does not mention this fact in the chapter, perhaps due to limited space.

nasalization has been lost' (240). Some languages have (apart from all the above) an impressive variety of verb suffixes; 'in some languages more than 130 verb suffixes are available' (244). The chapter ends by briefly describing a complex system of deictics.

Three chapters cover diminutive families: 'Makú' by Silvana and Valteir Martins (251-268); 'Nambiquara' by Ivan Lowe (269-292); and 'Arawá' by R.M.W. Dixon (293-306). Though the Makú family is now small, 'four languages belonging to seven tribes' (251); Makú speakers were the original inhabitants of the Brazilian Upper Rio Negro (and Colombian Vaupés) regions, later conquered and displaced by Arawakan and then Tucanoan groups. The phonology is quite distinct from surrounding languages, offering an abundance of CVC words (in a region where most languages are strongly CV oriented) and more complex vowel systems, except for Kakua, which seems to be heavily influenced by the surrounding Tucanoan languages.[10] Phonologies are actually more complex and fascinating than this summary would suggest: Yuhup has nasal harmony, phonetic pre- and post-nasalized voiced stops (Brandão Lopes and Parker 1999).[11] Daw, Kakua, and Yuhup have tone, in all cases incompletely analyzed.

Makú languages have much to offer the typologist: inalienable and alienable possession; locative postpositions which some might interpret as classifiers because 'their choice depends on the physical properties of the referent of the head' (258); noun incorporation (only in Nadeb); and ergativity (Nadeb, and possibly Kakua—the other languages are nominative-accusative). Within the Makú family, Nadeb stands out glaringly: this matter is discussed somewhat in Aikhenvald's chapter on areal diffusion in the Içana-Vaupés basin.

The tiny Nambiquara family (in Brazil) also has nasal prosody (not mentioned as such by the author), three contrastive tones, and contrastive laryngealization on nasal and oral vowels. Resyllabification of underlying consonant-final verb and noun roots occurs when these are suffixed. In the area of typology these languages offer well-developed evidential systems, exceeded in complexity only by those found in Eastern Tucanoan languages. There is also a complex system of verb subordination suffixes, noun classifiers (occurring on nouns, as 'deverbal nominalizers', modifying adjectives in an NP, and numerals), and a set of subject, object and copula

---

[10] There is an factual error in the text here (256, second line). Tucanoan languages have six vowel systems; Kakua does have a five vowel system, as the authors state.

[11] This reference was not available to the authors.

BEST COPY AVAILABLE

pronoun suffixes (including dual number). Lowe includes more on clause and interclausal syntax than do most authors in this book (277-279, 284-289); this has long been a special emphasis of his research.

The most notable characteristic of wider interest in the small Arawá family is probably the complex split ergative system, in which discourse topic, person, and noun suffixes all interact. The resulting construction then determines (for the most part) constituent order (299-300, 304-306). The languages have gender, and feminine is unmarked (masculine is considered to be marked). Some nouns 'require a cross-referencing prefix /ka-/ on the verb and on some nominal modifiers (when the noun is in pivot function in the clause)' (298). This is intriguing, in light of the /ka-/ of Arawak languages which has diffused into some Tucanoan languages; according to the analysis in Metzger 1999 one of the major functions of /ka-/ in Tucanoan and Arawakan is to indicate that a specific individual, item, or group is being focussed on. Some Arawakan languages are spoken in the general region where Arawá languages are found, though it does not look at first glance like much diffusion has taken place.

Two chapters discuss small language families and language isolates: 'Small language families and isolates in Peru' by Mary Ruth Wise (307-340); and 'Other small families and isolates' by Alexandra Y. Aikhenvald and R.M.W. Dixon (341-384). Both chapters represent a synthesis of the most disparate data to be treated in the book. Wise's discussion of Peruvian languages shows an amazing command of detail, in addition to skill at organizing what looks at first glance like hopeless disarray; it reflects long experience in the region (roughly 35 years) and previous practice in synthesizing large amounts of data (see Wise 1990 for just one nice example). With respect to phonology Wise observes that 'most of the languages in the five families differ from areal patterns in one or more traits' (312), all carefully listed (312-318). The most interesting to phonologists likely will be nasal prosody (here she refers to the published version of David Payne 1974), tonal systems (tone is not that common in the Amazon, as this book attests),[12] pitch accent systems in which stress (intensity) and high pitch (accent) do not necessarily coincide (described for Aguaruna in David Payne 1990a, but

[12] A fine resource now available for phonologists interested in tone systems of these languages (not available to Wise at the time of writing) is Walton et al. (1997); tone is marked on all entries in this dictionary. The third coauthor is a native speaker of Muinane, a member of the Witoto family.

according to Wise, probably typical of all Jivaro languages), and a nasal /h/ which nasalizes vowels that follow it (in Arabela).[13]

Most of the languages are rich in morphology. Characteristics of interest to typologists include classifiers, mostly numerical (one Witotan language, Bora, has over 350, and uses them with pronouns, in addition to the more usual nouns, demonstratives, adjectives and verbs). Dual number, unusual for the Amazon (as Wise notes) appear in the Witotan languages and Yagua; in Murui Witoto interaction with a three-way gender system results in a complex pronominal system. Some evidence suggests ergative traits in Jivaroan languages, but this possibility has not been fully explored. Verb morphology is especially complex, but the categories expressed are more or less typical for the Western Amazon.

Aikhenvald and Dixon's skills in dealing with complexity and disparity rival those of Wise; the wide distribution (Brazil, Bolivia, Colombia, and Venezuela) of small families and languages they summarize, plus a scarcity of sources, has made their job even more difficult. This variegated array of poorly described or basically undescribed languages certainly underscores the point made in the editors' introduction concerning the lack of attention given to Amazonian languages by the current linguistic world.

Languages covered which should be of interest to non-Amazonian linguists include the Yanomami dialect continuum (distinguished by 'a rich system of verbal classifiers' (347-348), multiple verbal proclitics and suffixes (over 20 on each end), extensive noun incorporation ('any noun in S or O function' (350)), and extremely productive verb compounding. Mura-Pirahã phonology is only briefly mentioned, but it is certainly one of the more unusual ones to be found in the book, with an unusually reduced consonant inventory (354) and a complex, as of yet only partially analyzed tonal system (according to Everett 1986).[14] The most well-known feature (among theoretical phonologists), that of syllable weight partially determined by consonant onsets, is not even mentioned here (see Everett and Everett 1984a,b).[15] Pirahã is also somewhat unusual that it lacks formal marking for tense, plurals, and possession.

---

[13] Walker and Pullum (1999) discuss this segment in a short report in Language, in which they argue that phonetically possible (or pronounceable) segments should not be excluded from phonology on theoretical grounds (see 769).

[14] To my knowledge, nothing has appeared since which would indicate otherwise.

[15] These references are not in the bibliography of this chapter. Readers who want to know more about Mura-Pirahã phonology and who do not read Portuguese should look at 308-325 of

The authors discuss the Guahiboan languages (Colombia and Venezuela) and some language isolates found in Colombia and Venezuela. Guahiboan languages are unusual in a number of areas: 'suppletive forms of verbs' depending on the number of the A or O (372); complex classifier systems (used with numerals, adjectives, and deictics) (373); 'an unusually large number of oblique cases compared to other Amazonian languages' (375); 'some traces of split ergativity, of an active-stative type;' incorporation of inalienably possessed nouns, either in S function with verbs of physical state, or S, O, or oblique function with other verbs.[16]

The last two chapters discuss two linguistic areas characterized by linguistic and cultural diffusion across several language families: 'Areal diffusion and language contact in the Içana-Vaupés basin, north-west Amazonia' by Alexandra Y. Aikhenvald (385-416); and 'The Upper Xingu as an incipient linguistic area' by Luci Seki (417-430). Both chapters will be of interest to sociolinguists and linguists specializing in language contact. Aikhenvald's chapter contains an impressive amount of linguistic detail, much of it based on her own fieldwork in both Arawakan and Tucanoan languages. The discussion covers most areas of phonology, morphology, and syntax; the author even includes a brief comparison of 'syntax and discourse techniques' (405-406). In contrast, Seki offers very little linguistic data; her intent apparently is to alert potential researchers to the possibility of rigorously documenting diffusional history from its beginning in a region where such work could be a real boon to struggling comparativists.

Diffusion in the Içana-Vaupés basin is most distinguished by the almost exceptional unidirectionality: the diffusion is from Eastern Tucanoan languages to geographically contiguous Arawakan and Makú languages. The only currently known exception is the Arawakan prefix /ka-/ 'relative, attributive' (Aikhenvald mentions this prefix in a footnote on p.392 and in her chapter on Arawakan languages—see p.95).[17] The unidirectionality is most certainly a reflection of sociolinguistic factors, as the author notes.

Everett (1986) (which is in the bibliography, and which does mention the role of consonant onsets in determining syllable weight).

[16] A couple of important references on the Guahibo language (Kondo 1985 and Queixalós 1985) do not seem to have been available to the authors, probably due to inaccessibility.

[17] A detailed description of the occurrence and function of /ka-/ in Tucanoan and Arawakan languages is found in Metzger (1998) (not available to Aikhenvald at the time of writing).

2.0 Comments on more general issues. Several issues raised by the editors' comments in the section on conventions and the book's introduction need more comment for the benefit of those who have not worked extensively in the Amazonian basin. I comment on these issues in the order in which they appear in the book, and I include an issue not covered which should be brought to the attention of those who wish to pursue field work in the Amazon or use scholarship from the Amazonian area (perhaps as a result of reading this book).

2.1 Terminology for language families. Specialists in South America when writing in English put an *–an* ending on the names of most language families. In Spanish, they do not, but instead write 'the family X', where X is the name without any adjectival (or other ending). In this book the editors follow the Spanish convention when writing in English. In many cases this means that a language family and a language in the same family will have the same name. They insist that context is enough to distinguish between the two, but my own experience suggests that this is not always the case.

In fact, I had never given the whole business much thought, until opening this book. When operating in Spanish, I (like most colleagues) automatically use the Spanish convention, and when using English, I follow the English convention. One must not only be bilingual with respect to vocabulary and grammar, but also with respect to conventions of language use. In this review article I have followed standard English convention, except when I quote from the book. It is noteworthy that in previous work and in at least one unpublished manuscript I have seen postdating the book the junior editor uses the standard English convention.[18]

2.2 Communication between missionary and Latin American linguists. Although the editors' observations are amazingly accurate regarding this topic, they are incomplete. After long years in Latin America, it appears to me that neither side has told the editors the whole story, perhaps because both sides know the editors communicate with anybody and everybody they can find working in the field, no matter what their religion, politics, or nationality (plus the junior editor, aside from being a polyglot, has an outstanding gift for relating to most anybody). For one thing, the Judeo-Christian ethics of most missionary linguists would preclude recounting to the editors negative incidents in a long history of ups and downs as they have sincerely tried to relate to local scholars and academic structures—all this in

---

[18] In Dixon (1994) the senior editor uses the *–an* ending and 'the X family.' I have not had access to anything he has written postdating this book, and am wondering if there might also be a difference between American and Australian usage here.

BEST COPY AVAILABLE

the face of concerted regional negative publicity, which extended from the seventies until the early nineties decade, and still (as the editors note) lingers in the region.

On the other hand, Latin American scholars are reluctant to comment on the highly politicized environment of too many universities, which has sometimes resulted in negative attitudes to the work of non-Latin linguists being a necessary part of maintaining one's job and position. In Latin America it is much harder to find jobs and funding that allow one to describe these languages; that does not help matters. Fortunately, the funding situation is beginning to change, due in part to the founding of several organizations devoted to the preservation of endangered languages, to the creation of regional networks such as GT Línguas Indígenas (devoted to the study of Brazilian languages) and to activities of organizations such as SSILA (Society for the Study of Indigenous Languages in the Americas).

Most important, in Latin American thinking linguistic work and associated literacy and development work is basically a political enterprise: it politically empowers minority peoples. The language family histories cited in the book (especially for the Tupí group and the Tupí-Guaraní subgroup) should make the implications of granting political power to minorities in the Latin American context clear. North American and European linguists for the most part do not look on linguistic and literacy work as political empowerment, but instead as part of meeting and respecting basic human needs and rights (along with access to medical services, clean water, and adequate nutrition), or as a basically apolitical good work. This divergence in world view has certainly made communication and mutual understanding between the two parties more difficult.

2.3. Definition of the Amazonian region. The editors state that 'in this book we attempt to cover languages spoken in the Amazon and Orinoco Basins—that is, from the north coast of South America, east to the mouth of the Amazon, west to the Andes, and south to the southernmost headwaters of the Amazon tributaries. If most of the languages in a family are spoken in the Amazon/Orinoco Basin (e.g. Arawak) then we cover that family. If most of the languages in a family are outside the region (e.g. Guaicuru) then we do not deal with that family' (4). This agrees with the definition proffered in the introduction to Derbyshire and Pullum (1986:1) and followed in Payne 1990. However, this definition excludes at least one language family in the northwest corner of the continent which seems more Amazonian than Andean: the Chocoan family, extending all along the Pacific coast and up into the western Andean range from eastern Panamá to the Colombian-Ecuadorian border. On the southern edge of Amazonia the Guaykuruan languages, according to Alejandra Vidal in her description of Pilagá

classifiers, present features typical of lowland Amazonian languages (Vidal 1997:102). This will be evident to readers of this book who go on to read her article in the now defunct Journal of Amazonian Languages.[19]

These families on the two extreme ends of the Amazonian region raise an important question: when we talk about Amazonian languages, are we defining a linguistic area, or a geographic area? This issue is still to be resolved, and any possible answers are complicated by preliminary, as of yet unpursued hints of linguistic diffusion between Andean and Amazonian languages.[20] It should be noted that linguists are not the only ones who have to cope with this problem: in a recently published field guide to neotropical rainforest mammals the author includes Central American Pacific coastal species that occur below 1000 meters altitude, observing that 'these species for the most part form a discrete fauna: mammal communities within the rainforest region are very similar to one another in their composition of monkeys, opossums, sloths, bats, deer, and rodents ... A few species known to occur only above 1,000 meters have also been included ... There are many borderline cases, and we have made some arbitrary decisions about which species to include' (Emmons 1997:1-2). This author clearly defines a biological area within a more general regional area, and there is also some residue.

2.4. Physical and logistical difficulties inherent in doing field work in the region. The editors do not comment on this matter, which is a shame, in light of the senior editor's impassioned plea in Dixon (1997) for more scholars to engage themselves in the documentation of undescribed, endangered minority languages. Those who are not familiar with the Amazon or South American region, should not sign up quickly to join the descriptive enterprise—it is wise to first count the cost, and then plunge in.

---

[19] A handful of isolates which might be potential candidates in Ecuador, Peru (see p.307) and Bolivia (see footnote 3 on p.364) are also excluded, but I have too little information to include them here.

[20] In a footnote Aikhenvald mentions the possibility of diffusion between Chocoan and Tucanoan languages with respect to nasal spreading. I was the source of this information; I have observed other shared characteristics (chiefly lexical vocabulary), but have not been able to study the similarities rigorously. In addition, she mentions the putative inclusion of the Chocoan family in the now generally discredited Macro-Chibchan subphylum in a footnote (370). The family has been considered an isolate, or alternatively stuck in the Carib family. A Chocoan linguist has commented to me that on the basis of his comparative studies he is tempted to locate the Chocoan languages in the Carib family. Until both the Carib and Chocoan families are studied more extensively, and more materials are available, it will be impossible to resolve these issues.

One pays in several areas: one is wear and tear induced by the physical difficulties inherent in working in the Amazon basin. Dan Everett provides a telling description in his review of Doris Payne's 1990 book (Everett 1991). All who have done long-term, serious fieldwork in the Amazon area have similar tales to tell; the details vary, but the overall theme is the same. More recently, in several countries, a more insidious danger has appeared: multiple armed illegal groups, mostly in marginal, isolated regions where a majority of poorly described minority languages are spoken. None of these armed groups care to see outsiders (no matter what nationality they may be) present on their turf, and usually employ brusque methods in dealing with this sort of perceived threat.

Another area in which one must be willing to deal is that of providing to the speech communities with which one works the results of research, in a form which they can understand and use for their own benefit. More and more speakers of minority languages are tired of researchers bopping in for a few months, doing research which demands extensive community participation, and disappearing forever, without any tangible benefits accruing to the people themselves. The result is that academic specialists are more and more likely these days to run into speakers who demand more than salaries for remunerating the language data that they share. These demands tend to crop up in less isolated groups, whose speakers are more aware of global trends, but more isolated groups may make other demands–usually not in the researcher's field of expertise. It takes time and money to meet these demands, no matter what their nature. No ethnic community that I know of shares the Western academic passion for pure research, none cares what their language will add to the Western European cultural heritage of acquired knowledge (why should they?). In their view, what researchers (and other outsiders) do in their community should be good for something practical. This again is a matter of world view, and it is sure to catch the unprepared scholar by surprise.

Another area in which one pays is that of professional advancement (see also the introduction to Payne 1990). It is difficult to conduct successful field work in South America without long term presence, or close association with scholars who have established the contacts and networks of relationships that come from long term residence in the area. Long term presence has its costs: field linguists resident in the Amazon area do not readily find the resources or time they need to keep up to date in linguistic theory, and the many unexpected demands associated with the descriptive task absorbs more time than any reader who has not worked in the area could ever dream possible. The result is frustrating: it is very hard to find the isolated blocks of time one needs to do heavy duty analysis and write theoretically oriented, data rich articles in one's native language—the kind of articles that major journals

might accept (forget purely descriptive articles—that hasn't been a hot trend for thirty years). Basically, this is the major factor which delays for so long the efforts of field linguists working in this area to cast research results in a form readily accessible to the outside world.

3. Conclusion. Non-specialists in South American languages perhaps have little idea of the courage and hard work that it took to produce this volume. Many of the reasons are touched on in their introduction. Nevertheless, the editors have succeeded in producing an accurate and ample overview of the current state of affairs in Amazonian linguistics, chiefly because they went to people who have been working for years in the area, and prevailed upon them to describe the languages and groups with which they are most familiar. One result is that six of the chapters are authored by SIL linguists and almost all of the other chapters lean heavily on material produced by or personal communication with SIL linguists. Four chapters are produced by Latin American linguists. Two (Ayron D. Rodrigues and Lucy Seki) have considerable experience in the area, whereas others are newer to the field (Silvana and Valteir Martins). This is a reflection of the incipient (and hopeful) trend of Latin American scholars to become involved in a field where there is much need for more hard descriptive and comparative work.

In fact, I think it is in part because Dixon and Aikhenvald are relatively new in the area (compared to most of the authors in the book) that they even dared to put together a book like this, and I'm frankly thankful they've pulled it off. As a result, we now have a useful, comprehensive linguistic anthology for the Amazonian region, something that did not exist before they stepped in.[21] Many of us who have worked longer in the area look more at what hasn't been done, then end up stalled by everything we can't do in the middle of an exceptionally difficult work environment. Some would say the enterprise is premature, but after reading the book I strongly disagree. My hope is that in company with Derbyshire and Pullum's work, this volume will be a potent force in changing the current descriptive situation for the better. Perhaps it will even have some influence on current linguistic theorizing, mostly taking place in parts of the world which seem like another planet as I sit in South America writing this review.

REFERENCES

Barnes, Janet. 1996. Autosegments with three-way lexical contrasts in Tuyuca. International Journal of American Linguistics 62.31-58.

---

[21] I must admit, I was using this book in my own field work—even before I had finished reading it!

Bendor-Samuel, J. T. 1960. Some problems of segmentation in the phonological analysis of Tereno. Word 16.348-355.

____. 1966. Some prosodic features in Terena. In Memory of J.R. Firth, ed. by C. E. Bazell, J. C. Catford, M. A. K. Halliday, and R. H. Robins, 30-39. London: Longmans.

Brandão Lopes, Aurise, and Steve Parker. 1999. Aspects of Yuhup phonology. International Journal of American Linguistics 65:324-342.

Cahill, Michael. 1999. From endangered to less endangered: Case histories from Brazil and Papua New Guinea. SIL Electronic Working Papers 1999-006. <www.sil.org>

Comrie, B. 1978. Ergativity. Syntactic typology, ed. by W.P. Lehmann, 329-94. Austin: University of Texas Press.

Craig, Colette, ed.. 1986. Noun classes and categorization. Amsterdam: John Benjamins.

Derbyshire, Desmond C. 1979. Hixkaryana. Lingua Descriptive Studies 1. Amsterdam: North Holland.

____. 1981. A diachronic explanation for the origin of OVS in some Carib languages. Journal of Linguistics 17.209-220.

____. 1983. Ergativity and transitivity in Paumari. Workpapers of the SIL and University of North Dakota 27.11-28.

____. 1985. Hixkaryana and linguistic typology. Dallas: SIL and the University of Texas at Arlington.

____. 1991. Are Cariban languages moving away from or towards ergative systems?. Workpapers of the SIL, University of North Dakota Session 35.1-39.

Derbyshire, Desmond C., and Geoffrey K. Pullum. 1981. Object initial languages. International Journal of American Linguistics 47.192-214.

Derbyshire, Desmond C., and Geoffrey K. Pullum, eds. 1986. Handbook of Amazonian languages: Volume 1. Amsterdam: Mouton de Gruyter.

____. 1990. Handbook of Amazonian languages: Volume 2. Amsterdam: Mouton de Gruyter.

____. 1991. Handbook of Amazonian languages: Volume 3. Amsterdam: Mouton de Gruyter.

____. 1998. Handbook of Amazonian languages: Volume 4. Amsterdam: Mouton de Gruyter.

Dixon, R.M.W. 1979. Ergativity. Language 55.59-138.

____. 1994. Ergativity. Cambridge: Cambridge University Press.

____. 1997. The rise and fall of languages. Cambridge: Cambridge University Press.

Emmons, Louise H. 1997. Neotropical rainforest mammals: A field guide, second edition. Illustrations by Francois Feer. Chicago: Chicago University Press.

Everett, Daniel. 1986. Pirahã. Handbook of Amazonian languages: Volume 1, ed. by Desmond C. Derbyshire and Geoffrey K. Pullum. 200-325. Amsterdam: Mouton de Gruyter.

____. 1991. Review of Amazonian linguistics: Studies in Lowland South American languages. Doris L. Payne, ed. Language 67.124-129.

Everett, Dan, and Keren Everett. 1984a. Syllable onsets and stress placement in Pirahã. West Coast Conference on Formal Linguistics 3.105-16.

____. 1984b. On the relevance of syllable onsets to stress placement. Linguistic Inquiry 15.705-711.

Gildea, Spike. 1995. A comparative description of syllable reduction in the Cariban language family. International Journal of American Linguistics 61.62-102.

____. 1998. On reconstructing grammar: Comparative Cariban morphosyntax. Anthropological Linguistics Series. Oxford: Oxford University Press.

Gómez-Imbert, Elsa. 1980. La nasalité en Tatuyo. Amerindia 5:65-81.

Harms, Phillip Lee. 1994. Epena Pedee syntax. Studies in the Languages of Colombia 4. Dallas: SIL and the University of Texas at Arlington.

Hayes, Bruce. 1995. A metrical theory of stress. Chicago: Chicago University Press.

Hyman, Larry M. 1982. The representation of nasality in Gokana. The structure of phonological representations, ed. by Harry van der Hulst and Norval Smith, 111-130. Dordrecht: Foris.

Kaye, Jonathan D. 1970. The Desano verb: Problems in semantics, syntax and phonology. New York: Columbia University dissertation.

Kaye, Jonathan D. 1971. Nasal harmony in Desano. Linguistic Inquiry 2.37-56.

Kondo, Riena de. 1985. *El guahibo hablado*. Vols. 1 and 2. Lomalinda, Colombia: Instituto Lingüístico de Verano.

Metzger, Ronald G. 1998. The morpheme KA- of Carapana (Tucanoan). SIL Electronic Working papers 1998-003. <www.sil.org>

Payne, David. 1974. Nasality in Aguaruna. Arlington, TX: University of Texas at Arlington thesis. (Published in 1976 as *Nasalidad en aguaruna. Serie Lingüística Peruana* 15. Pucallpa, Peru: Instituto Lingüístico de Verano.)

____. 1990a. Accent in Aguaruna. Amazonian linguistics, ed. by Doris Payne, 161-184. Austin: University of Texas Press.

____. 1990b. Some widespread grammatical forms in South American languages. Amazonian linguistics, ed. by Doris Payne, 77-87. Austin: University of Texas Press.

____. 1991. A classification of Maipuran (Arawakan) languages based on shared lexical retentions. In Derbyshire and Pullum, eds., 355-499.

Payne, Doris L. 1986. Noun classification in Yagua. Noun classes and categorization, ed. by Colette Craig, 113-131. Amsterdam: John Benjamins.

____. 1987. Noun classification in the Western Amazon. (Special issue: Comparative linguistics of South American Indian languages, ed. by Mary Ritchie Key.) Language Sciences 9.21-44.

____, ed. 1990. Amazonian linguistics. Austin: University of Texas Press.

Payne, Judith. 1990. Asheninca stress patterns. Amazonian linguistics, ed. by Doris Payne, 184-209. Austin: University of Texas Press.

Pike, Kenneth. 1964. Stress trains in Auca [Waorani]. In honour of Daniel Jones: Papers contributed on the occasion of his eightieth birthday, 12 September 1961, ed. by , D. Abercrombie, 425-431. London: Longmans, Green.

Pike, Kenneth, and Willard Kindberg. 1956. A problem in multiple stresses. Word 12.415-428.

Plank, L., ed. 1979. Ergativity: towards a theory of grammatical relations. London: Academic Press.

Quesada, J. Diego. 1999. Ergativity in Chibchan. Sprachtypol. Univ. Forsch. (STUF) 52:22-51.

Queixalós, Francisco. 1985. Fonología sikuani. Bogotá: Instituto Caro y Cuervo.

Safir, Ken. 1982. Nasal spreading in Capanahua. Linguistic Inquiry 13.689-694.

Smith, Richard, and Connie Smith. 1971. Southern Barasano phonemics. Linguistics 75.80-5.

Sparing-Chávez, Margarethe W. 1998. Interclausal reference in Amahuaca. Handbook of Amazonian Languages: Volume 4, ed. by Desmond C. Derbyshire and Geoffrey K. Pullum, 443-485. Amsterdam: Mouton de Gruyter.

van der Hulst, Harry, and Norval Smith. 1982. An overview of autosegmental and metrical phonology. The structure of phonological representations, ed. by Harry van der Hulst and Norval Smith, 1-46. Dordrecht: Foris.

Vidal, Alejandra. 1997. Noun classification in Pilagá (Guaykuruan). Journal of Amazonian Languages 1.58-111.

Walker, Rachel, and Geoffrey K. Pullum. 1999. Possible and impossible segments. Language 75.764-780.

Walton, James W., Janice P. Walton, and Clementina Pakky de Buenaventura. 1997. Diccionario bilingüe muinane-español español-muinane. Bogotá, Colombia: Editorial Alberto Lleras Camargo.

Waltz, Nathan, and Alva Wheeler. 1972. Proto Tucanoan. Comparative studies in Amerindian languages, ed. by Esther Matteson et al., 119-49. The Hague: Mouton.

Wise, Mary Ruth. 1990. Valence-changing affixes in Maipuran Arawakan languages. Amazonian Linguistics, ed. by Doris Payne, 89-116. Austin: University of Texas Press.

Terry Malone, A A1930, Santa Marta, Magdalena, Colombia, S.A.
Email: terry_malone@sil.org

■

Dissertation Abstracts

The interface between syntax and discourse in Korafe, a Papuan language of Papua New Guinea.

Published by **Pacific Linguistics** C-148.
Degree granted by The Australian National University 1996.

Cynthia J. M. Farr
SIL—Papua New Guinea

This dissertation focuses on the structure and function of three types of complex constructions which are central to Korafe discourse: (1) serial verb constructions (SVCs), (2) switch reference constructions (SRCs), and (3) co-ranking constructions.

Each of the complex construction types has as obligatory constituents two or more clauses or verbs. SVCs and SRCs are 'chaining' constructions, which terminate with a verb, more finitely inflected than the preceding verbal constituents, i.e. verb stems in SVCs and medial verb forms in SRCs. Syntactic constraints marked on or implicit in chaining structures enable the speaker to monitor subject reference from verb to verb without using many overt noun phrases. This reduction in noun phrases brings with it a corresponding focus on the events represented by the verbs in chaining structures. The order of the verbs in these chains is non-reversible and mirrors the order of the events they represent in the real world. This makes them choice vehicles for conveying the foregrounded story line in narratives, legends, and procedures. Utilising verbs without their standard arguments to a) represent familiar events (e.g. *ghambudo* 'dig' for 'dig a hole', *jedo govedo* 'chop plant' for 'making a garden'), and b) mark shifts in venue or temporal setting (*aira buvudo* 'he went and arrived', *ravara atetiri* 'they slept and it [day] dawned') enables the speaker to concentrate on the specifics of the story in question, using noun phrases to highlight dominant and/or prominent participants and props.

In 'co-ranking' constructions, all the constituents terminate with verbs of the same rank, namely final verbs, or in topic-comment constructions, predicate complements. Co-ranking constructions combine clauses, SRCs, and/or other co-ranking sentences by juxtaposing or conjoining them. Co-ranking constructions supply background information in discourses that primarily present events in iconic order. They are also extensively used in more thematically oriented discourses, such as encyclopedic descriptions, explanations, and hortatory speeches.

SRCs and co-ranking structures may be segmented into information chunks that are thematically unified. These thematic clause chain units (TCCUs) are defined by formal and semantic criteria. They range from one to nine words, comprising up to five clauses in an SRC and averaging between one to four seconds in length. They are uttered as a basically pause-free unit.

Cynthia J. M. Farr, SIL, Box 36, Ukarumpa, EHP 444, Papua New Guinea.
E-mail: jim-cindi_farr@sil.org

The dynamics of language spread: A study of the motivations and the social determinants of the spread of Sango in the Republic of Central Africa.

Degree granted by The Graduate Group in Sociolinguistics, University of Pennsylvania

Mark Edward Karan
SIL-Central African Group

Language spread, the process of a language expanding into new geographic and language-use areas, has been studied largely through observational methods. Thus discussions of the dynamics of the process have been largely based only on observation data. The present work employs a memory span test to evaluate the competence of a large number of subjects in a spreading language, Sango of the Republic of Central Africa. This large-sample, quantitative measure of competence enabled statistical studies of the social determinants and predictors of competence in the spreading language. The results indicate the overriding importance of motivations on the individual level in understanding the dynamics of language spread. Based on these motivations focused on the individual, a framework for discussion, research, and intervention in language spread is presented, along with guidelines for more successful intervention in shift situations. Numerous researchers have linked language spread and language change (language internal modification over time), but without substantive comparisons of the two. This quantitative study of language spread provides data on the distribution of social factors (age, sex, education, etc.). These distributions are very similar to the distributions of social factors in language change, indicating that language spread and language change are similar processes.

Mark Karan, SIL, B.P. 1990, Bangui, Central African Republic. E-mail: Mark_Karan@sil.org

■

Reports

Web-Based Language Documentation and Description and the Open Language Archives Community Workshop

J. Albert Bickford
SIL-Mexico Branch

This workshop (December 12–15, 2000, University of Pennsylvania) brought together linguists, archivists, software developers, publishers and funding agencies to discuss how best to publish information about language on the internet. The workshop and the Open Language Archives Community which is developing out of it seem especially important for us in SIL. I was pleased to be among those representing SIL, and hope that this report will be useful to others in SIL in understanding these new developments in the linguistics publishing and archiving field.

The aim of the workshop was to establish an infrastructure for electronic publishing that simultaneously addresses the needs of users (including scholars, language communities, and the general public), creators, archivists, software developers, and funding agencies. Such an infrastructure would ideally meet a number of requirements important to these different stakeholders, such as:

- Provide a single entry point on the internet through which all materials can be easily located, regardless of where they are stored (on the internet or in a traditional archive). Essentially, this would be a massive union catalog of the whole internet and beyond.
- Identify every language uniquely and precisely, so that all materials relevant to a particular language can be located.
- Make available software for creating, using, and archiving data (especially data in special formats); this includes software to help convert data from older formats to newer ones.
- Serve as a forum for giving and receiving advice about software, archiving practices, and related matters.
- Provide opportunity for comments and reviews of materials published within the system.

The workshop was organized by Steven Bird (University of Pennsylvania) and Gary Simons (SIL).[1] It included approximately 40 presentations and several working sessions on a variety of topics.

There was general agreement among the participants that a system for organizing the wealth of language-related material on the internet is needed, and that an appropriate way to establish one is by following the guidelines of the Open Archives Initiative (OAI) http://www.openarchives.org/. (These guidelines provide a general framework for creating systems like this for specific scholarly communities.) An OAI publishing and archiving system contains the following elements:

- Data providers, which house the materials that are indexed in the system.
- A standardized set of cataloguing information for describing each of the materials, also known as 'metadata' (i.e., data about data).
- Service providers, which collect the metadata from all the data providers and allow users to search it in various ways so as to locate materials of interest to them.

In the case of linguistics, the system will be known as the Open Language Archives Community (OLAC). The Linguist list http://www.linguistlist.org has agreed to serve the system as its primary service provider. It will be the main source that users will use to find materials through the system. Further information about OLAC can be found at http://www.language-archives.org. The agreement to establish OLAC is probably the most important accomplishment of the workshop.

This agreement was solidified through working sessions which met during the workshop and started the process of working through the details in various areas, such as:

- Character encoding: Unicode, fonts, character sets, etc.
- Data structure for different types of data (lexicons, annotated text, etc.).
- Metadata (cataloguing information that should be common to the whole community and how it should be represented in the computer) and other concerns of archivists.
- Ethics, especially the responsibilities that archivists and publishers have to language communities.

---

[1] Funding was provided by the Institute for Research in Cognitive Science (IRCS) of the University of Pennsylvania, the International Standards in Language Engineering Spoken Language Group (ISLE), and Talkbank.

- Expectations of users, creators (e.g. authors), software developers.

These and other issues will continue to be discussed on email lists in the coming months, ultimately culminating in recommendations for 'best practice' in each area, together with a preliminary launch of the whole system, hopefully within a year. (Prototypes of the system are available now at the gateway address above, along with various planning documents.)

There were also a number of conference papers which provided a foundation for making the working sessions productive. Rather than list or review all the presentations here, I will summarize them, since they are all available on the conference website

http://www.ldc.upenn.edu/exploration/expl2000.

The topics covered included the following:

- Proposals for various aspects of the OLAC system.
- Concerns of various stakeholders, such as archivists, sponsors, language communities.
- Descriptions and demonstrations of specific software, research projects, and web publishing systems.
- Metadata and metadata standards.
- Technical issues, such as Unicode, the OAI, sorting, data formats for different types of language materials (e.g. dictionaries, annotated text, example sentences in linguistic papers, and audio).

One insight gleaned from these presentations was a better understanding of glossed interlinear text. Interlinear text is not a type of data, but rather just one possible way of displaying an annotated text. The annotations on a text can consist of many types of information: alternate transcriptions, morpheme glosses, word glosses, free translations, syntactic structure (and possibly several alternative tree structures for the same text), discourse structure, audio and video recordings, footnotes and commentary on various issues, etc. What ties them all together is a 'timeline' that proceeds from the beginning to the end of a text, to which different types of information are anchored. Aligned interlinear glosses are one way of displaying some of this information, but not the only way, and not even the most appropriate way for some types of information. The traditional arrangement of Talmudic material, for example, with the core text in the center of the page and commentary around the edges, is another possible display of annotated text, in which the annotations are associated more with whole sentences and paragraphs than with individual morphemes. There are also some sophisticated examples available for presenting audio alongside interlinear text. See a sample at LACITO archive as follows:

http://lacito.archive.vjf.cnrs.fr!)

Throughout, it was very clear that those at the conference had a great deal in common with each other, and there are many points of contact with our concerns within SIL:

- A primary concern for descriptive (as distinguished from theoretical) linguistics.
- A desire to make language materials available to communities of speakers and the general public as well as scholars.
- An interest in taking advantage of the Internet, which provides a means of publishing such materials that by-passes the limitations of traditional publication (since the costs are so much lower, and thus appropriate for materials that have smaller audiences) .
- Awareness that many materials may be less than fully-polished yet still valuable to some people and worth archiving .
- A sense of frustration with the currently confused state of the art in data formats, especially fonts and character encoding, and the lack of good information about how best to archive and publish on the web.
- Awareness of the large amount of data that is in data formats which will be obsolete in a few years (and thus a willingness to accept data in whatever form it is in, while also seeing a need for software to help convert data to newer formats).
- A strong suspicion toward and distrust of rigid requirements, yet a willingness to adopt standards voluntarily when their usefulness has been demonstrated.

Given that we have so much in common, it was very appropriate that SIL was so actively involved in the workshop and that we should continue our involvement.

- Several SIL members will be active in the continuing discussions leading toward inauguration of the OLAC system.
- Ethnologue codes will almost certainly be used as unique language identifiers within the system, and SIL will be recognized as the 'naming authority' for assigning them. (Other naming authorities will hopefully also be recognized, particularly for areas that the Ethnologue does not provide codes for, such as ancient languages and language families.)

BEST COPY AVAILABLE

- SIL software is of interest and use to other people, especially products like Graphite (for supporting complex non-Roman scripts) and Shoebox. In our software development efforts we would do well to keep in mind the needs of linguists outside of SIL. Their needs are not very different from ours, and sometimes small changes in software design can make big differences to eventual usability (for us as well as them).
- Other people are developing specialized linguistic software that could be very useful to us or which contains ideas which we can incorporate in our own software, especially including tools for dictionaries and annotated text.
- There are many people (linguists, archivists, and software developers) who are more than happy to share their expertise with us as we work through our own archiving and publication issues.
- There is funding available to help us develop our own archives, provided we are willing to share the contents with others.

In short, this conference opened up avenues of cooperation with a number of important partners in the academic world—a class of potential partners that has tended to be ignored in discussions within SIL in recent years.

Finally, the conference pointed out several trends that will be increasingly important in future years, both inside and outside SIL.

- The speakers of lesser-known languages will be more actively involved in the production and use of materials in and about their languages, and their concerns will increasingly have to be considered by scholars. These include carefully documenting permissions and levels of access to materials, making sure that language materials are available to the communities themselves, and being careful that scholars do not inadvertently aid commercial interests in exploiting native knowledge-systems (such as medicinal use of plants) without appropriate compensation.
- The boundary between publishing, libraries, and archiving is being blurred by the shift to the digital world. Materials can be 'archived' on the web, which is a type of publication. Electronic 'libraries' are springing up in many places. Published and unpublished works from around the world can be listed together in one common catalog. The same technology is important in both spheres of activity. In short, these activities are merging under a new umbrella that could be called 'scholarly information management'. A corollary to this trend is that archiving is not just something you

do at the end of a language project; it is part of the ongoing process of managing the information that the project produces.

- In such a world, and with huge numbers of resources available to sift through, metadata becomes increasingly important. A freeform paragraph description in a publications catalog is no longer good enough. It is the metadata that users will consult in order to find materials of interest to them, so the metadata must be carefully structured, accurate, and current. More and more, we will have to think not just about producing materials but also about how to describe them so as to make them accessible to others.
- Unicode http://www.unicode.org is the way of the future for representation of special characters in computers. The days of special fonts for each language project are numbered. Instead, Unicode will make possible a single set of fonts that meets virtually everyone's needs in the same package. Over the next few years all of us will be switching our computers over to using Unicode almost exclusively (that is, if we want to take advantage of newer software). Our computer support personnel are already actively involved in this transition, and have been for several years. It has already impacted in various ways on users within SIL, even if we haven't realized it, but from now on the impact will be much more apparent. Get ready to change your fonts one more time—but once you've done so, that should be the last.
- Language data will increasingly need to be structured carefully so that not only can people view it and use it, but machines will be able to understand and manipulate it in various ways. Many of us are familiar with standard format markers, which have been the primary means within SIL of marking the structure of data for over 15 years. Despite its usefulness, it has some limitations. Standard format will, within the next few years, be replaced by a more comprehensive system called XML which is widely-supported in the computer industry and which can do everything that standard format can do and more.[2]

---

[2] 'XML' stands for 'Extensible markup language'. Since its development has been closely-associated with the World Wide Web consortium http://www.w3.org/XML/, it has been widely regarded as the successor to HTML for web pages. However, this is just a small part of its usefulness; it is a general-purpose system for representing the structure of information in a document or database, which can be customized for myriad of purposes. Many software tools are currently available for creating and manipulating data in XML, with more being created all the time. One, Extensible Style-sheet Language Transformations http://www.w3.org/TR/xslt, can do complex restructuring of XML data, including many of the things that we have used CC for in the past, but with far less programmer time.

BEST COPY AVAILABLE

All in all, it was a workshop that was both stimulating and practical—one which will have an unusual amount of influence in months and years to come.

J. Albert Bickford, P. O. Box 8987, Catalina, AZ 85738-0987. E-mail: albert_bickford@sil.org

International Linguistic Association, 46th Annual Conference. Languages of the Americas, Native and Non-native. New York, NY. March 30, 31, April 1, 2001

TERRY MALONE
SIL—Colombia

According to their website (http://www.ilaword.org) The International Linguistic Association was founded in 1943 in part at the impulse of exiled European scholars, and the membership still is international. Although in the two decades after its founding the ILA played a significant role in the world of American theoretical linguistics, the advent of Chomskyan linguistics has done much to overshadow and even eclipse it. This seems a shame: most members are dedicated to advancing theoretical linguistics by practicing applied linguistics. Few ILA members subscribe to the Chomskyan paradigms, but this is perhaps the society for those scholars who need a variety of contacts yet can't afford to belong to half a dozen specialized societies all at once. Fellow members will include text linguists, typologists, Indoeuropean linguists, systemic linguistics, specialists in TESOL, university teachers of foreign languages, bilingual educators, sociolinguists, and others outside the current American theoretical mainstream. All are doggedly practicing linguistics as the society's founders conceived of the discipline; hopefully some day the fashionable tides on American linguistic shores will turn, so that these scholars will receive a fraction of the appreciation that they deserve.

The conference theme resulted in about a day's equivalent of papers and plenary sessions on Native American languages, and another day's equivalent of papers on Spanish, Portuguese, and AAVE (African American Vernacular English), plus a panel on the latter. There was about a half-day's equivalent of papers on more general linguistic themes. Scholars such as Aaron Broadman, Jill Brody, Harriet Klein, Marianne Mithun, David Payne, and Rachelle Waksler presented papers or conducted plenary sessions on Native American languages, along with less well-known scholars, linguistic

'grunts' (i.e. OWLS such as myself), and a handful of scholars who do not specialize in Native American linguistics but who drew instructive typological parallels with more well-studied languages.

The papers that I was able to hear on Native American languages all had immediate relevance to current theoretical issues, both within and outside of the Chomskyan mainstream. These included hot topics such as morphemes which do not quite fit in any of the categories 'clitic', 'affix' or 'word' ('Floating morphemes in an Amazonian language' by Sidney Facundes), discourse markers which elude simple analysis at any grammatical level ('Marking Focus in Chatino' by Troi Carlton and Rachelle Waksler), grammatical categories whose membership is not clean-cut ('The category of adjectives in Southern Guaykuruan languages' by Alejandra Vidal and Harriet E. Manelis Klein), gender systems ('Cross-linguistic view of Gender in Ojibwe' by Donald Steinmetz), and multiple classifier systems in one language ('Classifiers in Chimila' by yours truly).

The three plenary sessions were highpoints for me: David Payne discussed 'areas of linguistic typology for which South American languages challenge current linguistic wisdom', along with some possible reasons for such typological divergence; Ofelia García discussed a hypothesis she is exploring concerning what happens when speakers of a minority language (Spanish in New York City) try to shift to a majority language (English) but do not receive the rewards for language shift that they had anticipated; and Marianne Mithun illustrated the crucial need to take diachronic processes into account when explaining typological correlations (especially those which appear to violate typological universals).

The small size of the conference (roughly 60 papers were listed in the handbook and a few of these cancelled) made it possible for me to interact personally with many of the scholars specializing in Native American linguistics—that is perhaps what I most appreciated about this conference. They all had feedback to offer this linguistic 'grunt', and that ardently coveted feedback was a major reason I responded to Ruth Brend's and Mike Cahill's initial call for papers. My biggest disappointment was to see that other than David and Judy Payne, I was the only SIL member from the Americas to attend this conference. SIL linguistic 'grunts' have much to offer the members of ILA, and they have much to offer us; both parties suffer when such opportunities are ignored.

Terry Malone, AA 1930 Santa Marta, Magdalena, Colombia, S.A.
E-mail: terry_malone@sil.org

■

Reviews

Chomsky's universal grammar, 2nd edition. By VIVIAN COOK and MARK NEWSOM. Cambridge, MA: Blackwell Publishers, Inc. 1996. 369 pp. Hardback $62.95, paper $25.95.

Reviewed by ALAN BUSEMAN
SIL—ICTS, Waxhaw

Noam Chomsky has had a profound impact on modern linguistics. This book is an extremely readable introduction to his theories. The first edition of this book—published in 1988—covered the theory known as Government and Binding. This second edition is revised to cover more recent versions of Chomsky's theory including Principles and Parameters Theory and the Minimalist Programme.

The authors state their purpose on page. 1:

> The aim of this book is to convey why Chomsky's theory of language is stimulating and adventurous and why it has important consequences for all those working with language ... This book is intended chiefly as an introduction for those who want to have a broad overview of the theory with sufficient detail to see how its main concepts work, rather than for those who are specialist students of syntax ...

The authors fulfill their purpose extremely well. The book is easy to read and very interesting. The examples are fun. The presentation is nicely done—with key concepts summarized in boxes with a gray background. The editing is extremely good—with almost no noticeable errors.

This book is a perfect starting point for a linguist from a different theoretical framework who wants to get an overview of Chomsky's theories. It introduces the basic concepts and defines the basic terminology in a very thorough and inviting way. The text also includes hundreds of quotes and references which will be helpful to anyone who wants to read further on the subject.

Why would you as a field linguist want to read this book? One reason is that Chomsky's concepts are so pervasive in linguistics that you need to understand them and know the vocabulary used to express them to help you understand much of the other current literature in linguistics. Another reason is that since Chomsky's grammar is universal, it should apply to the language you are studying. This may help you start with a more effective set

BEST COPY AVAILABLE

of expectations and can lead to a number of interesting questions about the language. You may even find the language you are studying relates in some interesting way to the proposed universal grammar, and you may be able to publish something about it.

This is one of my favorite kinds of books and I thoroughly enjoyed it. It looks back enough to show the most influential writings and lines of thought in the development of the theory, and it is current enough to summarize the main current lines of thought. It helped me understand Chomsky's motivations and reasoning.

If you wish you knew more about what Chomsky has been up to and why his influence in linguistics has been so powerful, start with this book.

Alan Buseman, P. O. Box 248, Waxhaw, NC 28173. E-mail: Alan_Buseman@sil.org

Meaning in language: An introduction to semantics and pragmatics. By D. ALAN CRUSE. New York: Oxford University Press. 2000. 424 pp. Paper $24.95

Reviewed by GEORGE HUTTAR
SIL—Africa Group

Here is a valuable reference work for your library. Its coverage in semantics, especially lexical semantics, is broad and detailed, and it includes a useful introduction to pragmatics as well. A summary of its organization is probably as good a way as any to give you an idea of what it covers, though this minimal mentioning of topics falls far short of doing justice to the quality of C's coverage:

Part 1 'Fundamental Notions' opens, as you might expect, with Ch. 1 'Introduction' (3-16). Situating the study of meaning within linguistics and other disciplines, C also makes clear what approach to expect throughout the book: e.g., 'Meanings are not finitely describable, so [the] task [of specifying or describing meanings] boils down to finding the best way to approximate meanings as closely as is necessary for current purposes' (13). While the book takes an 'ecumenical' position as far as theories go ... (14):

> ... in so far as there is a theoretical bias, it is towards the cognitive semantic position. This means, in particular, that the meaning of a linguistic expression is taken to arise from the fact that the latter gives access to a particular conceptual content. This may be of indeterminate extent: no distinction is made between linguistic meaning and encyclopedic knowledge.

BEST COPY AVAILABLE

Ch. 2 'Logical matters' (17-39) covers arguments and predicates; sense, denotation, and reference (including intension and extension); sentence, statement, utterance, and proposition; logical classes; logical relations' quantification; and use and mention—all those fundamental notions you have to keep straight when thinking through seriously what's going on in basic sentence and lexical semantics.

C's contextual approach to meaning shows up again in Ch. 3 'Types and dimensions of meaning' (41-63), where meaning is characterized as 'anything that affects the relative normality of grammatical expressions' (43)—a characterization which leads to a brief but useful treatment of how to distinguish grammatical from semantic anomaly. Most of the chapter describes dimensions of descriptive meaning—the objective meaning that determines truth value and reference—and of non-descriptive meaning. To the former belong differences of quality (*red* vs. *green*, *dog* vs. *cat*), intensity (*large* vs. *huge, scared* vs. *terrified*), specificity (*dog* vs. *animal*), vagueness, and six other dimensions. The latter covers expressive meaning (*Stop blubbering* vs. *Stop crying*) and variants according to register (further broken down into field, mode, and style) and dialect. This number and variety of distinctions is typical of the whole book's degree of detail. Part 1 concludes with Ch. 4, 'Compositionality' (65-81), the focus of which is 'on the way meanings combine together to form more complex meanings'.

Part 2 'Words and Their Meanings' is introduced thus (with page numbers added) (83):

> To the layman, words are *par excellence* the bearers of meaning in language. While it is in danger of understating the importance of other linguistic structures and phenomena in the elaboration of meaning, this view is not entirely unjustified: words do have a central role to play in the coding of meaning, and are responsible for much of the richness and subtlety of messages conveyed linguistically. Hence it is no accident that this part of the book is the most substantial. Here, after the introductory Chapter 5 [85-102], we discuss how word meanings vary with context (Chapter 6 [103-124]), the relations between word meanings and concepts (Chapter 7 [125-142]), paradigmatic sense relations (Chapters 8 [143-161] and 9 [163-176]), larger vocabulary structures (Chapter 10 [177-196]), how new meanings grow out of old ones (Chapter 11 [197-216]), how words affect the meanings of their syntagmatic neighbours (Chapter 12 [217-235]), and finally, theories of lexical decomposition (Chapter 13 [237-261]).

Its introductory chapter usefully summarizes six basic approaches to semantics, besides outlining the basic problems of lexical semantics and explicating the distinction between lexical and grammatical meaning, the relation between word meaning and sentence meaning, and the interesting

BEST COPY AVAILABLE

question of what are, and what are not, possible meanings of words in any language (while some languages have a single word (or stem) for a meaning like DRINK SLOWLY—cf. Eng. *sip*—none is known to have one for WOMAN DRANK or THE WINE SLOWLY). A sample of topics from the rest of Part 2 includes: ambiguity; homonymy and polysemy; classical and prototype approaches to categorization; basic-level categories; hyponymy, meronymy, and synonymy; various kinds of opposites; taxonomic and meronymic hierarchies; bipolar chains (e.g., *minuscule, tiny, small, large, huge, gigantic*); metaphor; metonymy; semantic change; types of combinatory abnormality (e.g., collocational clash); componential analysis, and semantic primitives.

Part 3 'Semantics and Grammar' consists of Chapter 24 'Grammatical semantics' (265-300); is survey of 'those aspects of the meanings of larger syntactic units which are attributable to grammar' (263); includes detailed description of the grammatical meanings associated with basic lexical classes. The discussion of number on nouns, for example, mentions singular, plural, dual, trial, and paucal; count nouns and mass nouns; basic count nouns used as mass nouns; basic mass nouns used as count nouns; the semi-mass use of count nouns; singular nouns with (optional) plural concord; plural nouns with (optional) singular concord. The discussion of aspect deals not only with distinctions like perfective/imperfective, perfect/prospective (the latter has to do with the present relevance of a future event), punctual/durative, punctual/iterative, and inchoative/medial/terminative, but also with the aspectual character of specific verb stems which affect their ability to occur with various aspectual markers, etc.

Part 4 'Pragmatics' consists of three chapters: Ch. 15 'Reference and deixis' (305-327), Ch. 16 'Speech acts' (329-346), and Ch. 17 'Implicatures' (347-378). C's approach here can be inferred from a list of the scholars whose work is most frequently referred to: Grice, Searle, Leech, and the relevance theorists Sperber, Wilson, and Blakemore.

An unnumbered chapter, 'Conclusion' (379-381) rounds out the book with a list of what C sees as the six major areas of uncertainty about fundamental issues in meaning in language.

The usefulness of this book is enhanced by the following features: a bibliography of around 150 items; 'Suggestions for further reading' at the end of each chapter; 'Discussion questions and exercises' at the end of almost every chapter with answers to most of the questions in a separate section; a subject index, and an author index.

As a textbook, this is not for your beginning students (as C himself makes clear). And it is not always easy for more advanced readers, especially those not at home in British English or, in some cases, English culture. For a textbook more accessible to students from a variety of national educational systems, try Saeed 1997, but get Cruse's book for a thorough reference work on a number of semantic and pragmatic topics, especially in lexical semantics. Its combination of breadth and detail offer a useful array of stimulating ideas to explore with regard to any language in its own right, as well as for purposes of application to effective translating.

REFERENCE

Saeed, John I. 1997. Semantics. Oxford: Blackwell Publishers Ltd.

George Huttar, Box 24686, Nairobi, Kenya. E-mail: george_huttar@sil.org

Foundations of statistical natural language processing. By CHRISTOPHER D. MANNING and HINRICH SCHÜTZE. Cambridge, MA: The MIT Press. 1999. 680 pp. $60.

Reviewed by MIKE MAXWELL
SIL—International Administration

Statistical approaches to linguistic analysis predate generative linguistics—indeed, Noam Chomsky wrote the epitaph of statistical linguistics in the 1950s. Why, then, is this textbook about statistical natural language processing (NLP) coming out now? A complete answer would be too large for this review; suffice to say that whatever the scientific reasons for preferring generative linguistics (or a competing theory), when it comes to implementing practical applications on computers, engineering takes the forefront. And for better or worse, engineers do not always imitate nature. This, then, is a book about engineering approaches to language, and specifically those approaches that generate probabilistic answers.

While acknowledging that engineering approaches to NLP are indispensable in certain contexts, I feel compelled to devote a few paragraphs at the beginning of this review to questioning some of Manning and Schütze's more far-reaching claims, as they impinge on the science of linguistics. The following quote (taken from the beginning of a chapter on acquiring lexical patterns from large corpora) is one such claim (p. 265):

> While we discuss simply the ability of computers to learn lexical information from online texts, rather than in any way attempting to model human language

> acquisition, to the extent that such methods are successful, they tend to undermine the classical Chomskyan arguments for an innate language faculty based on the perceived poverty of the stimulus.

There is much that could be said about this. The learning of lexical information has seldom been the focus of arguments for an innate language faculty, and certainly not Chomsky's focus; the fact that some generalizations may be drawn from large corpora does not negate arguments based on the poverty of stimulus, since there are well-known generalizations for which one is unlikely to find sufficient supporting data in corpora of any size. But rather than argue these points, I will observe that the above statement can be turned on its head to the extent that statistical methods by themselves do not succeed in learning linguistic generalizations, they reinforce the classical arguments for an innate language faculty. Time and more research in statistical methods will tell how true this is—but there are hints even now.[1] At the end of the same chapter from which the above quote was taken, the authors write (p. 311):

> What does the future hold for lexical acquisition? One important trend is to look harder for sources of prior knowledge that can constrain the process of lexical acquisition. This is in contrast to earlier work that tried to start 'from scratch' and favored deriving everything from the corpus ... One important source of prior knowledge should be linguistic theory, which has been surprisingly underutilized in Statistical NLP.

Having stated most of my negatives up front, I hasten to add that regardless of whether statistical NLP has anything to say about the innate vs. learned knowledge debate, there are many practical applications where statistical NLP shines. In field linguistics, for example, some potential applications include determining similarity among related languages/dialects; probabilistic disambiguation of parsed text (such as choosing the best morphological parse of a word, based on the words in its environment); preliminary categorization of collocations and word senses; and part of

---

[1]For those interested in this issue, I will give one other quote, from a discussion on Markov chains (p. 378):

> Chomsky's criticism still applies: Markov chains cannot fully model natural language ... What has changed is that approaches that emphasize technical goals such as solving a particular task have become acceptable even if they are not founded on a theory that fully explains language as a cognitive phenomenon.

This emphasis on engineering, not science, is prevalent throughout the text, and quite understandable given the goals of NLP. But it makes the authors' comments quoted in the text to the effect that the results of statistical NLP 'undermine' arguments for an innate language faculty seem rather curious.

speech tagging for purposes of syntactic analysis—not to mention potential applications to literacy and translation.[2]

One word of caution, though: there is no magic bullet. Statistical NLP tends to require large corpora. Just what 'large' means—and how that might vary depending on such language-particular factors as amount of inflectional morphology—is for the most part glossed over in this book.[3] What the reader is occasionally told may prove discouraging to field linguists: the English text of Tom Sawyer, with 71,370 word tokens, is 'a very small corpus by any standards, just big enough to illustrate a few basic points' (p. 21). It takes a long time to collect 70,000 words of text in a preliterate village.

Foundations of Statistical Natural Language Processing is intended as a graduate level text, but is equally suited to the linguist who wants to know what all the fuss about statistical NLP is. One application that is not covered is that of speech recognition, which has been especially indebted to statistical approaches—including it would have made a large book considerably larger. There are a number of other topics to which the authors say they have not given in-depth coverage, including 'machine learning, text categorization, information retrieval, and cognitive science' (p. xxxi). Actually, a chapter each is devoted to information retrieval and text categorization, and the chapters on clustering and text categorization are especially indebted to machine learning techniques (in supervised and unsupervised classification).

Following an introduction to the goals of statistical NLP, there are introductory chapters on probability theory, mathematical information theory, and linguistics (the latter with an emphasis on structures encountered in English). The reader is well advised to have some prior background in probability and statistics. My sense as a linguist is that in this book explanations of linguistic matters are often simplistic. This is not a condemnation; giving more attention to classical linguistics would only have made an already formidable volume even larger.

---

[2]I emphasize the word 'preliminary'. Manning and Schütze write (p. 18):

> It is interesting to look at the types of collocations that a purely linguistic analysis of text will discover if plenty of time and person power is available ... a wider variety of grammatical patterns is considered [in the manually compiled dictionary entries they show] ... Naturally, the quality of collocations is also higher than computer-generated lists—as we would expect from a manually produced compilation.

[3]Unfortunately, most statistical NLP seems to have been done on English, a language notorious for its lack of inflectional morphology.

Next follows an eminently practical chapter on the issues of working with language corpora, including the problems theoretical linguists gloss over: punctuation, upper/lower case, word and sentence division, and text markup. (For markup, Manning and Schütze recommend SGML, but add that XML—a slimmed-down version of SGML—is likely to take over for purposes of NLP. My sense as I write this review is that this has already happened.) Also discussed are several tagging schemes for English; while these are referred to in later examples, the tags are too English-specific to be useful for other languages.

The remainder of the book is divided into three sections. The first of these is 'Words', which begins with a chapter concerning methods for discovering collocations, nicely illustrated by an example discriminating the meanings of the English adjectives 'strong' and 'powerful' based on with which nouns each word co-occurs. This sort of application should be easy to add to a concordancing program, although its usefulness will depend on how often the words in question appear in one's texts. (Manning and Schütze's example used a corpus of 14 million words, about 115 megabytes.)

Other chapters in this section discuss discovering word senses and lexical acquisition in general. A chapter on dealing with sparse data is also included, presumably because it can be illustrated by problems dealing with words. The authors briefly discuss a technique needed for languages with a great deal more inflectional affixation than English, namely stemming (removing inflectional affixes). One wonders whether certain purposes might not be better served by 'affixing', that is, ignoring the stems of inflected words.

The section on 'Grammar' includes Markov models which are probabilistic finite state automatons. Markov models are often useful for predicting the next event based on the last few events; in a linguistic context, 'events' might be words or parts of speech. For certain kinds of Markov models, 'grammars' can be learned automatically by a computer from a corpus. While these are not what a linguist would think of as the grammar of a natural language, this may not matter for some applications. One such application is further developed in the next chapter on part-of-speech (POS) tagging. Since investigators working on POS tagging have worked mostly on English, the techniques have emphasized syntagmatic context and lexical information, largely ignoring inflectional affixation. This has two implications for tagging in other languages. First, POS tagging should work well for other languages which largely lack inflection. Secondly, tagging may prove useful in more highly inflected languages for disambiguating morphologically derived POS information, although to the extent that such a

BEST COPY AVAILABLE

language has 'free' word order, syntagmatically-based POS tagging may not work well (as Manning and Schütze observe).

The grammar section is rounded out by chapters on probabilistic context-free grammars (phrase structure grammars with rules annotated for their probability) and probabilistic parsing (which involves assigning a probability to a particular structural analysis given its context, one application of which is disambiguating syntactic parses probabilistically).

The section entitled 'Applications and Techniques' includes a chapter on the alignment of bilingual corpora, a technique which has been used to create bilingual dictionaries and parallel grammars. Attempts have also been made to use the results of such alignment directly for statistically-based machine translation, but Manning and Schütze suggest that these efforts have failed because they incorporate too little linguistic knowledge. Alignment is most easily done with literal translations of texts such as legal documents; Manning and Schütze state that since 'religious and literary works' tend not to be translated so literally, they are more difficult to align.[4] Structural differences between the languages whose texts are to be aligned naturally add to the difficulty, although the authors cite work that has been done to overcome this.

A chapter on clustering compares algorithms for finding similarities among objects in general, and words in particular. An example would be grouping words into classes which might correspond to parts of speech. The authors note that 'The efficiency of clustering algorithms is becoming more important as text collections and NLP data sets increase in size' (p. 527), which implies that these techniques may prove less useful in field linguistics.

Finally, there are two chapters on information retrieval and text categorization, applications of obvious importance in this day of the Internet.

For the most part, the text seems admirably clear; charts, graphs and formulas are well laid-out, and mathematical derivations are straightforward. Technical terms are indicated in the margins at their first appearance, which together with a passable index makes it easy to look up a term you missed the first time through. (This reviewer confesses to having needed that help on more than one occasion.) However, while proper names are well represented in the index, topics are not as well indexed. For example, there is one entry starting with COMPUT ('computational lexicography'), which is

[4]On the other hand, scripture, as well as some literature such as plays, often comes pre-aligned at roughly the sentence level which should make word-level alignment easier.

BEST COPY AVAILABLE

either one too many (since nearly every section talks about computation), or far too few; there is no entry for 'programming language' (or anything similar), although the topic is discussed on page 121.

Scattered throughout the text are exercises, some of the pencil-and-paper variety, some which require the use of publicly available programs, and some which require actual programming.

The authors (and the publisher) are to be commended for setting up a companion web site (HYPERLINK http://nlp.stanford.edu/fsnlp) containing (or pointing to) the complete table of contents and two sample chapters; corpora; executable programs; and—not least—errata. The latter is sizeable, but given the technical nature of this book, and the large number of mathematical formulae, I would hesitate to call it unreasonable. (Some of the errata were corrected in the second printing.)

Mike Maxwell, 7809 Radin Rd., Waxhaw, NC 28173. E-mail: Mike_Maxwell@sil.org

Demonstratives: Form, function and grammaticalization. By HOLGER DIESSEL. Typological Studies in Language, 42. Amsterdam: John Benjamins. 1999.

Reviewed by DAVID MEAD
SIL—Indonesia Branch

Ask a linguist how much they know about demonstratives and they might answer: some languages have 'person-oriented' systems (having a special term for something near the addressee) while other languages have 'distance oriented' systems; languages differ in regard to how many distance-terms they distinguish; some languages distinguish elevation in their demonstrative system, others don't, etc.

According to Holger Diessel, factors such as these are important for understanding the semantics of demonstratives, but semantics is only one aspect. There are, in fact, three other facets of demonstratives which must be grasped in order to understand and fully describe these forms, either cross-linguistically or in a particular language.

Following are four different facets. At the outset I found Holger Diessel's book to be insightful and well organized and give it high recommendations, with only a few criticisms mentioned below.

Foundational to Diessel's approach to demonstratives is his insistence that we clearly distinguish the syntactic positions (distributions) in which

BEST COPY AVAILABLE

demonstratives may occur. Minimally there is a need to recognize (a) adnominal demonstratives which appear with a co-occurring noun (pop THAT balloon); (b) pronominal demonstratives which occur independently in argument positions of verbs and adpositions (hit THAT again); (c) adverbial demonstratives which are clausal level modifiers (leave it THERE); and (d) identificational demonstratives, which occur in copular and non-verbal clauses (THAT is my cousin). An important reason to keep these distinctions clear is that some languages (such as English) may use the same form in two or more of these syntactic positions, while other languages may formally distinguish demonstratives in all four positions. Padoe, an Austronesian language of Indonesia, is of the latter type. Diessel introduces four corresponding category labels, which may appropriately be used of Padoe, since this language has four formally distinct categories of demonstratives (Padoe data are from my own sources):

no-langkai	*mia*	**la'a**	'that person is big'	
he-big	person	that	(demonstrative determiner)	
no-tarima-'o	*kee*	**ula'a**	'did he receive that?'	
he-receive-it	INTERROG	that	(demonstrative pronoun)	
tila'a	*inehu*	*henu*	*winawa-nggu*	'those are the vegetables I brought'
that	vegetable	REL	brought-my	(demonstrative identifier)
m-powangu	*raha*	**lehea**	'(they) built a house there'	
PLURAL-build	house	there	(demonstrative adverb)	

(Two additional categories which languages may formally distinguish are manner demonstratives such as Padoe *helela'a* 'like that', and deictic presentatives such as French *voilà*.) English on the other hand, notes Diessel (p. 79):

> ... does not distinguish between demonstrative pronouns and demonstrative identifiers. The demonstratives in copular clauses have the same phonological and morphological features as pronominal demonstratives in other contexts and hence they are considered demonstrative pronouns.

BEST COPY AVAILABLE

Once this basic distinction between distribution and category is grasped, the rest of the book makes for easy reading.

The book itself is organized into seven chapters. Chapter 1, the introduction, lays the groundwork for Diessel's study and introduces the distinction between distribution versus category. Chapter 2 is a look in detail at the demonstrative systems of four different languages, giving the reader an expectation of how widely languages can differ in this area.

The next four chapters constitute the heart of the book, and each in turn deals with one of the four facets important for fully understanding demonstrative systems. Chapter 3 outlines and gives examples of the SEMANTIC distinctions which may be encoded in demonstrative systems (e.g. distance, visibility, elevation, animacy, number, etc.). Chapter 4 considers the SYNTAX of demonstratives, wherein Diessel revisits the distribution versus category distinction in much greater detail. Chapter 5 looks at the PRAGMATIC uses of demonstratives—that is, their primary, exophoric use in referring to objects present in the speech situation, and various endophoric uses such as when demonstratives are used in referring to participants or propositions in surrounding discourse. Chapter 6 looks at the GRAMMATICALIZATION of demonstratives. Here again one must pay attention to distribution: only PRONOMINAL demonstratives, for example, are likely to become third person pronouns, while only ADNOMINAL demonstratives are likely to grammaticize to become definite articles. Finally the entire contents of the book are summarized in the concluding Chapter 7.

Diessel is to be commended for including data from a wealth of languages. Ideas are explained clearly, and the organization of the book is such that—once the basic principles are grasped—one could probably read the remaining chapters in any order. Field workers at a beginning stage of analyzing a language are likely to find chapters 3 and 4 to be the most useful. The remaining chapters also have their place, and in fact chapter 6 on grammaticalization may hold the key for explaining 'odd' behavior of certain demonstratives (to wit, they are on their way to becoming some other grammatical category).

I do have a few criticisms of the book. One is that Diessel addresses the issue of demonstrative directionals only incidentally. He notes, for example, that one semantic category that demonstratives may encode is direction (toward speaker, away from speaker, across field of vision of speaker), but curiously the English forms *hither*, *thither*, *hence* and *thence* never come up for discussion (later, however, the German directionals *hin*—and *her*—are given as examples of demonstratives which have grammaticalized as

preverbs). This left me at a loss as to how to incorporate certain languages of Indonesia into his system which—in addition to having all four categories of demonstratives given above—have yet another category of demonstratives which imply motion in a particular direction, e.g. Padoe *ramai* 'there (coming toward here)', apparently distributing as a verb and as a noun modifier.

Regarding his category of demonstrative identifiers, Diessel considers both French *c'* (as in *c' est Pascal* 'this is Pascal') and Ponapean *iet* (as in *iet noumw pinselen* 'here is your pencil') to represent this category. But as the English glosses indicate (*c'* being translated as 'this', *iet* as 'here'), one wonders if this collapsing is really justified.

In addition, there is surely more to discover regarding the grammaticalization of demonstratives, and Diessel's chapter 6 cannot yet be the complete story. In Padoe alone one may note the grammaticalization of the deictic presentative *nio* (be.here-it) as an existential particle, and the grammaticalization of the demonstrative determiner *sie* 'that (near)' as a subordinate clause marker (*ro-me-hawe sie...* /they-PLURAL-arrive that/ 'upon their arrival ...'). Neither of these possibilities are mentioned by Diessel. By the way, his table on page 155, titled The Grammaticalization of Demonstratives, lists more possibilities than are actually described in the accompanying text. One wonders if this could have resulted from editorial changes to the text that were not incorporated into the table.

The reader should also be aware that Diessel uses the term 'expletive' in its secondary, less encountered sense which he defines as 'semantically empty pro-forms that some languages require to form certain syntactic constructions' (p. 149)—compare *there* in *there was nothing left*. Since cognitive linguistics studies have cast doubt on the entire concept of semantically empty pro-forms, Diessel's use of the term is doubly regrettable.

Donald Mead, P. O. Box 81439, 8000 Davao City, Philippines. E-mail: don_mead@sil.org

The Cambridge Dictionary of American English.
By SIDNEY I. LANDAU, ed. New York: Cambridge University Press.
2000. 1,069 pp. Paper $15.95, CD-ROM $20.95.

Reviewed by JOHN E. STARK
University of Ilorin, Linguistics Dept.

The renowned fictional detective, Nero Wolfe, once sat by the fireplace in his front room burning the third edition of Webster's New International

Dictionary page by page (Stout 1962, 1967). Wolfe, a connoisseur (a person who has expert knowledge of something, especially in art, food, or drink, and is qualified to judge and appreciate its quality (Landau 2000) of food and words) deemed it 'subversive and intolerably offensive'. I have long admired Wolfe's command of the English language, and was curious to know why he felt that way. A few pages later in the novel, Archie Goodwin, Wolfe's assistant, asked a client 'Do you use imply and infer interchangeably?' Webster's 3rd allowed it, and Wolfe took offense. A few years back I found a copy of Webster's 3rd, and read the front matter. It stated the definitions chosen were determined by common usage. Wolfe, in 1963, was functioning as a prescriptivist (prescriptive, adj tending to say what someone should do or how something should be done. His basic attitude toward language was prescriptive (Landau 2000)).

Unless his linguistic reading has converted Wolfe to a descriptive mindset over the intervening decades, he would also burn the Cambridge Dictionary of American English. Its preface states: 'This dictionary describes the ways American English is used now ...' It is based on a 100 million word corpus of written and spoken English taken from recent books, magazines, newspapers and broadcast sources which allows it to 'give examples that are based on genuine, present-day writing and speech (Landau 2000).

In using the Cambridge Dictionary for three months, I found I always agreed with the definitions of words I thought I knew. That makes me like it. Yet words I had not heard before were not included (e.g. limned, as in 'She wanted no transcendent beauty of scene, no terrible convulsion of nature, limned on her memory' (Grey 1935, 1991)). In addition, it lacks any type of synonym, cross-reference or thesaurus material. As a reference source for a person holding a bachelor's degree from an American University, it's 40,000 words and phrases are surprisingly inadequate, but then, such a person is not likely the target audience.

The work was prepared with consultation from ESL/EFL experts around the world, and utilizes a limited defining vocabulary of 2,000 words. It includes usage labels such as 'rude slang' or 'taboo slang'. These items indicate a target audience of non-English speakers, people without college education, and classroom use for high school and lower grades. Nonetheless, it has features of interest to every speaker of English.

One special feature is a series of boxes imbedded in the text at the location of a key word for a functional feature of English, called Language Portraits. On the page spread for 'sight—silently' is a box labeled 'Silent Letters' with examples and 'rules where possible' for when a, b, c, d, e, g, h, k, l, n, p, s, t, and w are silent in written English. Another feature I found helpful was the

BEST COPY AVAILABLE

inclusion of 'key words' behind the entry to indicate the nature of the following definition: e.g. mania [strong interest] / mania [mental illness], each followed by an appropriate expansion of the meaning involved.

This copy came with a software version on CD-ROM. It contains the full body of the dictionary, and an audible pronunciation feature that does a good job of producing a standard American English pronunciation of the word in isolation. However, I have used other computer dictionaries that had a more satisfying user interface (the connection between a person and a computer (Landau 2000), without a specific complaint against the Cambridge software it just felt cheesy (adj cheap or of low quality) (Landau 2000)).

This dictionary has elements which would be useful to incorporate into mother-tongue dictionaries: the arrangement of head-words and subentries, the quality of the illustrative sentences, and the Language Portraits would all bear examination as models for other dictionaries. It would also be a very helpful tool for second-language English speakers, and I expect that it will quickly become the most popular reference dictionary in the Kambari Language Project office, where a full-time staff of eight people, from three Nigerian languages, work to produce literacy and translation materials.

That the Cambridge Dictionary example sentence describes Wolfe so aptly is an example of the success of a descriptive approach. It receives a 'should buy' rating.

REFERENCES

Grey, Z. 1935, 1991. Thunder Mountain. New York, Harper.

Landau, S. I., ed. 2000. The Cambridge Dictionary of American English.. New York: Cambridge University Press.

Stout, R. 1962, 1967. Gambit. Bantam.

John E. Stark, E-mail: John_Stark@sil.org:

Direct reference, indexicality, and propositional attitudes. By WOLFGANG KÜNNE, ALBERT NEWEN, and MARTIN ANDUSCHUS, eds. CSLI Lecture Notes No. 70. Stanford, CA. New York: Cambridge University Press. 1997. 402 pp. Hardback $69.95, paper $24.95.

Reviewed by CARL WHITEHEAD
SIL-Papua New Guinea

Most of this volume consists of revised versions of papers presented at a conference of the same name held at Bielefeld, Germany in March 1994.

BEST COPY AVAILABLE

(Two of the 18 papers were submitted later.) The conference brought together linguists and philosophers working in the domain of possible world semantics, specifically within the framework of David Kaplan's *Demonstratives, An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals,* which was eventually published in 1989 (*Themes from Kaplan* edited by J. Almog, J. Perry and H. Wettstein. New York: Oxford University Press, pp. 481-563). The articles are by no means introductory and one author states explicitly: 'I have to assume in this paper that the reader is acquainted with Kaplan's semantical framework' (Nida-Rümelin p. 385). Several of the papers make use of semantic formalism which can be confusing to the uninitiated. The volume is divided into Indexicals and names (eight papers), Attitude reports (five papers), and Natural kind terms and color terms (three papers). For the sake of readers who are not familiar with the term indexicals, in the pure sense they are terms which refer directly to the utterance context such as *I, here* and *now*.

I found the volume useful for giving insight into another realm of linguistic semantics and for bringing the realisation that even the use of pronouns and proper names is semantically far more complex than we normally assume. Yet for most SIL translator/linguists, not having the background knowledge of possible world semantics and the work of Kaplan specifically, the insights gained from the articles would not warrant the time invested in understanding them.

What follows is a summary of the articles in the first section of the book. This should not be taken as an indication that this section is more accessible or valuable than the others. It is simply that, having skimmed the collection and then read this section in sufficient detail to provide the following summaries, considerations of time and length of the report dictated this limitation.

In the opening article, 'Reflexivity, indexicality and names', John Perry seeks to explain the difference in cognitive significance between 'I am a computer scientist' and 'David Israel is a computer scientist—both said by David Israel. This is accomplished by acknowledging the difference between reflexive and incremental content. The former of these refers to truth conditions which are about the utterance itself and is present in the first sentence to a greater degree than in the second. Indexicals and (proper) names have in common that they refer rather than describe (that is, they contribute the entire object they designate to the content); they differ in that whereas indexicals specify a condition the referent must meet, names designate their referent directly.

In 'Tensed thoughts' James Higginbotham also appeals to reflexivity but to differentiate between tensed and tenseless thoughts. Tensed thoughts predicate a state which exists at the time of the thought. Beliefs, wishes, feelings of regret, relief, or anticipation 'can only be directed toward thoughts that are themselves tensed or else supported by tensed thoughts, which locate the time reference of the untensed thoughts with respect to the thinker's present state' (p 24).

The following three articles all deal specifically with first-person reference (*I*). Wolfgang Künne ('First person propositions: A Fregean account') asks: 'What is the Fregean *sense* [=mode of presentation] which completes the sense of the predication "has blood type A" in ["I have blood type A"]?' and answers it from within the notions proposed by Frege around the turn of the century. He proposes a formalization of the 'ego mode of presentation' which is extendable to other indexicals. While working within Fregean philosophy, he does disagree with Frege on some points, including 'the idea that successful communication culminates in content sharing' (p 65). *I* is a hybrid proper name with each user forming a different TYPE expression referring to a proper name, and the speaker will never be able to successfully communicate the same content for *I* that he/she has in mind. Using substitute indexical counterparts, however, the addressee can express the same thought as the original.

Christopher Peacocke ('First-person reference, representational independence, and self-knowledge') differentiates between representational dependent uses of *I* and representational independent uses. In the former (72):

> ...the thinker is in a state which represents the content C as correct, and the content represented as correct is one which stands in such a justificational relation to the content of the belief 'I am F'.

Representational independent uses, such as *I see the phone is on the table*, are the result of a transition from the occurrence of a conscious event or state to a self ascription. He challenges Kant and others' assertion that such uses of *I* do not refer but have a transcendental or metaphysical subject. He uses the dependent/independent distinction to explain how such sentences do have subjects which do refer.

Albert Newen ('The logic of indexical thoughts and the metaphysics of the 'self'') proposes and defends a new logical formalization for thoughts ***de se*** (about oneself). He summarizes the attributes which such a formal representation would need to meet and demonstrates how his proposal best does so. As with Künne, the proposal includes a special ego mode of presentation, to 'represent the immediate way in which I am related to myself while having a

BEST COPY AVAILABLE

thought ***de se***' (118). He adopts in part an earlier proposal of Peacocke's, distinguishing between a type mode of presentation and a token mode thereof. Under the new proposal, the type mode of presentation (MP) is indexed by the context to yield a token mode. The formalism can be generalized to apply not only to ***de se*** thoughts but also to ***de re*** (about other things) thoughts. The generalized version is $<[MP_{context}];$object *o*> which would, for Mach's thought *I am a shabby pedagogue*, would be filled out as: $(<[ego_{context}];$Mach>;being a shabby pedagogue).

Thomas Zimmermann's 'The addressing puzzle' is the only article which focuses on the second person indexicals, arguing that the difference between formal and informal pronouns (primarily German *Sie* and *du*) is pragmatic rather than semantic (truth-condition-related). On a more general note, he also argues that 'No two sentences are uttered in precisely the same (possible) situations and hence they cannot have the same truth conditions'. The Most Certain Principle is, therefore, replaced with the Least Certain Principle: 'If S and S' are sentences, then S and S' differ in truth conditional meaning'.

The final two papers in the section reject Frege's proposal of senses or modes of presentation as being unnecessary. Henk Zeevat, in 'The mechanics of the counterpart relation', presents a model of how meaning (objects and ideas) are communicated from one mind to another. Counterparts can be created within the mind either of real external items or, via communication, of items internal to the speaker's mind. Reference to non-existent objects can then be explained as reference to items that are only internal to the speaker's mind, whereas Frege claimed that mention of non-existent objects was not reference at all. The nature of presupposition and assertion is discussed in detail, including the accumulative nature of presupposition and the process of finding a referent with the body of presupposition resulting in either resolution (finding the referent) or accommodation (adding a new item). Ernesto Napoli's 'Names, indexicals, and identity statements' addresses the nature of proper names and indexicals and argues that both are meaningless free variables which only refer in the context of an utterance. Frege's puzzle as to how an identity statement containing two proper names (e.g. Tom is Dick) can be informative whereas identity statements of the form $a=a$ are analytically true and therefore not informative, is answered by challenging those evaluations of informativeness. Since proper names are not constants, 'an identity statement of the form 'a =a' can be false and a fortiori not a priori true' (192).

Carl Whitehead, P. O. Box 4997, Three Hill, Alberta T0M 2N0 Canada.
E-mail: whitehead@will.kneehill.com

Bilingual acquisition: Theoretical implications of a case study. By MARGARET DEUCHAR and SUZANNE QUAY. 2000. New York: Oxford University Press. 163 pp. Cloth $70.00.

Reviewed by CATHERINE YOUNG
SIL—Philippines Branch

Bilingual Acquisition presents the findings of a case study in bilingual language acquisition and explores the implications of these findings for theories of first and second language acquisition. The research that forms the basis for this book began as an ESRC and British academy funded project entitled 'Infant Bilingualism: One System or Two?' Research grants enabled Margaret Deuchar to collect audiovideo data from her developing English-Spanish bilingual daughter between the ages of 1:3 and 3:3. The primary aims of this book are to explore the implications for linguistic theory of a case study in bilingual acquisition and involves the discussion of a number of general theoretical questions.

- Can phonological distinctions be acquired on acoustic evidence alone?
- Does lexical development involve an avoidance of synonymy?
- Can all words in early two-word utterances be assigned to lexical categories?
- How do young children make appropriate language choices?

Specific implications for bilingual acquisition include the following questions:

- Does the bilingual child have one or two linguistic systems from the beginning?
- What criteria should be used in identifying one versus two systems?
- What are the most important determinants of language choice for the developing bilingual?

Significant description of the methodology involved in eliciting the data for the study, a combination of journal and audiovisual records, is included and would be very useful for anyone intending to plan similar case studies.

Bilingual Acquisition contains extensive appendices reflecting the child's cumulative lexicon and multiword utterances.

Catherine Young; P. O. Box 2270, CPO; 1099 Manila, Philippines.
E-mail: catherine_young@sil.org

■

Books Available for Review

The following new books are available for review. Contact the editor and the book will be shipped to you along with instructions for submitting the review. When you submit a review, the book will be yours to keep. Contact: Mike Cahill, Editor, *Notes on Linguistics*, 7500 W. Camp Wisdom Road, Dallas, TX 75236 or Internet e-mail: Eugenecc_Loos@sil.org (Editorial Assistant)

Bauer, Laurie. 2001. Morphological productivity. New York: Cambridge University Press. 245 pp. Hardback $59.95.

Berg, Thomas. 2001. Linguistic structure and change: An explanation from language processing. New York: Oxford University Press. 236 pp. $27.95.

Bybee, Joan. 2001. Phonology and language use. New York: Cambridge University Press. 247 pp. Hardback $59.95.

Cowie, A. P., ed. 2001. Phraseology: Theory, analysis, and applications. New York: Oxford University Press. 258 pp. Cloth $24.95.

Deutscher, Guy. 2001. Syntactic change in Akkadian. New York: Oxford University Press. 204 pp. Cloth $70.00.

King, Jeffrey C. 2001. Complex demonstratives: A quanificational account. Cambridge: The MIT Press. 207 pp. Cloth $45.00, paper $18.00.

Krishnamurti, Bhadriraju. 2001. Comparative Dravidian linguistics. New York: Oxford University Press. 417 pp. $95.00.

Marantz, Alec, et al, eds. 2001. Image, language, brain. Papers from the First Mind Articulation Project Symposium. Cambridge, MA: The MIT Press. 272 pp. Cloth $55.00.

Matthews, Peter. 2001. A short history of structural linguistics. New York: Cambridge University Press. 163 pp. Hardback $59.95, paper $21.95.

Melamed, I. Dan. 2001. Empirical methods for exploiting parallel texts. Cambridge, MA: The MIT Press. 195 pp. Hardcover $32.00.

Moro, Andrea. 2001. Dynamic antisymmetry. Cambridge: The MIT Press. 142 pp. Paper $20.

Muysken, Pieter. 2001. Bilingual speech: A typology of code-mixing. New York: Cambridge University Press. 306 pp. Hardback $59.95.

Nation, I.S.P. 2001. Learning vocabulary in another language. New York: Cambridge University Press. 477 pp. Hardback $74.95, paper $29.95.

Neidle, Carol, et al. 2000. The syntax of American Sign Language: Functional categories and hierarchical structure. Cambridge: The MIT Press. 229 pp. $35.00.

Newman, Paul and Martha Ratliff, eds. 2001. Linguistic fieldwork. Cambridge University Press. 288 pp. Hardback $69.95, paper $24.95.

Palmer, F. R. 2001. Mood and modality. New York: Cambridge University Press. 236 pp. Hardback $64.95, paper $22.95.

Pesetsky, David. 2001. Phrasal movement and its kin. Cambridge: The MIT Press. 132 pp. Paper $18.00.

Pintzuk, Susan, George Tsoulas, and Anthony Warner. 2001. Diachronic syntax: Models and mechanisms. New York: Oxford University Press. 380 pp. Paper $39.95.

Siptár and Miklós Törkenczy. 2001. The phonology of Hungarian. New York: Oxford University Press. 319 pp. Hardback $99.00.

Stump, Gregory T. 2001. Inflectional morphology: A theory of paradigm structure. New York: Cambridge University Press. 308 pp. Hardback $64.95.

Van Valin, Jr., Robert D. 2001. An introduction to syntax. New York: Cambridge University Press. 239 pp. Hardback $59.95, paper $22.95.

■

BEST COPY AVAILABLE

CONTENTS
(Continued)

REVIEWS

CHOMSKY'S UNIVERSAL GRAMMAR by Vivian Cook and Mark Newsom — Alan Buseman 60

MEANING IN LANGUAGE: An introduction to semantics and pragmatics by D. Alan Cruse — George Huttar 61

FOUNDATIONS OF STATISTICAL NATURAL LANGUAGE PROCESSING by Christopher D. Manning and Hinrich Schütze — Mike Maxwell 64

DEMONSTRATIVES: FORM, FUNCTION AND GRAMMATICALIZATION by Holger Diessel — David Mead 69

THE CAMBRIDGE DICTIONARY OF AMERICAN ENGLISH by Sidney I. Landau, ed. — John E. Stark 72

DIRECT REFERENCE, INDEXICALITY, AND PROPOSITIONAL ATTITUDES by Wolfgang Künne, Albert Newen and Martin Anduschus, eds. — Carl Whitehead 74

BILINGUAL ACQUISITION: THEORETICAL IMPLICATIONS OF A CASE STUDY by Margaret Deuchar and Suzanne Quay — Catherine Young 78

ANNOUNCEMENTS

BOOKS AVAILABLE FOR REVIEW 79

BEST COPY AVAILABLE

NOTES ON *LINGUISTICS*

VOLUME 4, NUMBER 2 2001

CONTENTS

COORDINATOR'S CORNER

THE VALUE OF HISTORICAL AND COMPARATIVE LINGUISTICS Michael Cahill 83

ARTICLES

THE VALUE OF COMPARATIVE LINGUISTICS Joseph E. Grimes 85

HISTORICAL LINGUISTICS IN SOUTHEAST ASIAN LANGUAGE PROGRAMS Paulette Hopple 93

COMPARATIVE-HISTORICAL MESOAMERICAN RECONSTRUCTION AND SIL PERSONNEL: ACCOMPLISHMENTS AND PROBLEMS Robert E. Longacre 106

REPORTS

THE 28TH LACUS FORUM Shin Ja Hwang 111

LACUS 2001: THE NATURE OF LINGUISTIC EVIDENCE Robert E. Longacre 115

REVIEWS

POSSESSIVES IN ENGLISH: AN EXPLORATION IN COGNITIVE GRAMMAR by John R. Taylor Chaoqun Xie 120

ENGLISH TRANSCRIPTION COURSE by M. Luisa Garcia Lecumberri and John A. Maidment Karen Steffen Chung 124

THE TURKISH LANGUAGE REFORM: A CATASTROPHIC SUCCESS by Geoffrey Lewis Alan S. Kaye 126

LEARNER-CENTREDNESS AS LANGUAGE EDUCATION by Ian Tudor Mary Morgan 129

(Continued on back cover)

SIL INTERNATIONAL
7500 W. CAMP WISDOM ROAD
DALLAS, TEXAS 75236 USA

+L 027421

NOTES ON LINGUISTICS

INTERNATIONAL LINGUISTICS COORDINATOR/EDITOR ■
Michael Cahill
ASSOCIATE EDITOR ■ *Eugene Loos*
CONSULTING EDITOR ■ *Alan S. Kaye*
FORMAT EDITOR / SECRETARY ■ *Betty Philpott*

INTERNATIONAL LINGUISTICS ADVISORS

Thomas Bearth	*U. of Zurich*
Steven Bird	*U. of Pennsylvania*
Ruth Brend	*Michigan State U., r.*
Derek Fivaz	*Rhodes U., r.*
Peter Fries	*Central Michigan U.*
Shirley Brice Heath	*Stanford U.*
Yoshihiko Ikegami	*U. of Tokyo*
Frances Ingemann	*U. of Kansas*
Peter Ladefoged	*U.C.L.A.*
Sydney Lamb	*Rice U.*
W. P. Lehmann	*U. of Texas at Austin*
Howard P. McKaughan	*U. of Hawaii, r.*
Dinh Hoa Nguyen	*Southern Illinois U., r.*
John Oller	*U. of Southwestern Louisiana*
A. K. Pawley	*Australian National U.*
Richard Rhodes	*U. of California, Berkeley*
Malcolm Ross	*Australian National U.*
William J. Samarin	*U. of Toronto*
Hu T'an	*N.R.C. on Tibetology, Beijing*
Rudolph Troike	*U. of Arizona*
Richard Tucker	*Carnegie Mellen U.*
Robert Van Valin	*SUNY, Buffalo*
Ekkehard Wolff	*U. of Leipzig*

INTERNATIONAL LINGUISTICS CONSULTANTS

Rudy Barlaan	*Philippines*	Lon Diehl	*Asia*	Steve Marlett	*Mexico*
John Bendor-Samuel	*U.K.*	Warren Glover	*South Asia*	Kenneth McElhanon	*Dallas*
Andy Black	*Mexico*	Kenneth J. Gregerson	*Asia*	Doris Payne	*U. of Oregon*
Cheri Black	*Mexico*	Joseph Grimes	*Hawaii*	Tom Payne	*U. of Oregon*
Regina Blass	*Burkina Faso/Niger*	Austin Hale	*South Asia*	Randy Radney	*Canada*
Les Bruce	*Dallas*	Lou Hohulin	*Dallas*	Gary Simons	*Dallas*
Donald A. Burquest	*U.T., Arlington*	George Huttar	*Nairobi*	Keith Snider	*Canada*
Gene Casad	*Mexico*	Connie Kutsch Lojenga	*Africa*	David Weber	*Peru*
Rod Casali	*Africa*	Stephen H. Levinsohn	*Columbia*	Klaus Wedekind	*East Africa*
Ross Caughley	*Asia*	Ivan Lowe	*U.K.*	Ursula Wiesemann	*Germany*

SIL International President ■ *Carolyn Miller*
SIL International Executive Director ■ *John Watters*
SIL International VP for Academic Affairs ■ *Karl Franklin*

NOTES FOR READERS ■ Permission is hereby granted by this journal to reproduce material from this issue, if readers obtain similar permission directly from its authors and if credit is given the authors and *Notes on Linguistics*.

NOTES FOR SUBSCRIBERS ■ *Notes on Linguistics* is a quarterly publication of the Linguistics Department of SIL International, designed to share information of practical, theoretical, administrative, or general interest. It is specifically intended to provide linguistic field workers with news, reviews, announcements, and articles that will stimulate interest in linguistics and help them stay current.

SUBSCRIPTION ■ Annual subscription rate: U.S. address (delivered by U.S. Postal Service) $20.95. Outside U.S. (delivered by AIR) $23.95. A 20% discount is offered to resellers and personnel of SIL International. Prices include postage and handling. Subscriptions will be prorated so that all subscriptions end with the fourth issue of the volume.

INDIVIDUAL ISSUES ■ Individual copies of available issues are $3.00 each, plus a minimum of $2.75 handling (which generally covers about 3 issues). Not all back issues are still available.

ORDERING INFORMATION Prepayment is requested from individuals; reseller and institutional purchase orders are accepted. Payment must be in U.S. funds, and checks must be drawn on a U.S. bank. Major credit cards are welcomed; to pay by credit card please include account number and expiration date. Electronic transfers are not accepted. Subscriptions, standing orders, and back issues are handled by: International Academic Bookstore, 7500 W. Camp Wisdom Road, Dallas, TX 75236, USA. Phone: 972-708-7404 FAX: 972-708-7363; Email: academic_books@sil.org

NOTES FOR CONTRIBUTORS ■ Readers are invited to submit publishable materials or letters of comment to the editor of *Notes on Linguistics* at The International Linguistics Center, 7500 W. Camp Wisdom Road, Dallas, TX 75236 USA. Internet e-mail: Mike_Cahill@sil.org (articles, etc.) or Eugene_Loos@sil.org (reviews). Intending contributors should ensure that their manuscripts conform to the standards laid down in the LSA Style Sheet printed periodically in the *LSA Bulletin*. Contributors who have copies of their manuscripts on magnetic media such as diskettes in PC/MS-DOS are urged to submit the same along with their manuscript. (3½" PC/MS-DOS diskettes or as e-mail attachments, in Word for Windows format are preferred.)

DISCLAIMER: Responsibility for the accuracy of quoted and referenced material rests solely upon the author of the item in which such material appears. Opinions expressed by the authors do not necessarily reflect policies of SIL.

ISSN 0736-0673
BEST COPY AVAILABLE

COORDINATOR'S CORNER

The value of historical and comparative linguistics

Historical linguistics got its main impetus in 1786, when Sir William Jones wrote:

> The Sanscrit language, whatever be its antiquity, is of a wonderful structure; more perfect than the Greek, more copious than the Latin, and more exquisitely refined than either, yet bearing to both of them a stronger affinity, both in the roots of verbs and in the forms of grammar, than could possibly have been produced by accident; so strong indeed, that no philologer could examine them all three, without believing them to have sprung from some common source, which perhaps no longer exists...

In 'examining them all three', tone would find such similarities as the word for 'two', which is Greek *duō*, Latin *duo*, and Sanskrit *dvau*. The root for 'foot' is Greek *pad-*, Latin *ped-*, and Sanskrit *pad-* (Hock and Joseph 1996). Jones' claim set off a flurry of research. Eventually the ancestral language of Greek, Latin, and Sanskrit (now spelled with a *k*) was reconstructed, labeled Proto-Indo-European, and is still an area of intense historical research.

Language changes! Below is the first sentence of the 'Lord's Prayer' in Old English, around 1100 A.D., along with the King James Version of 1611:

> Fæder ure þu þe eart on heofonum, si þin nama gehalgod.
> Our father who art in heaven, hallowed be thy name.

Syntax, semantics, sounds, orthography—all change. We see many of these illustrated in the 500 year span shown above. (The 'þ' above is a voiceless interdental fricative called 'thorn', and 'æ' has its modern IPA value).

The theme of this issue of *Notes on Linguistics* is historical and comparative linguistics. One might think that these are not SIL's concern; we are more concerned with modern languages. But historically (!) SIL has contributed much to these areas, as in the reconstruction of Proto-Otomanguean, at a time depth comparable to Proto-Indo-European. David Thomas and others have contributed considerably to Mon-Khmer studies. SIL's on-line bibliography reveals 584 entries listed under 'historical' and 'comparative'. Why this emphasis?

Historical linguistics can give you an explanation for an otherwise puzzling language pattern. In Kɔnni (Cahill 2000/2001), many nouns have a stem-final /g/ which deletes before the singular suffix *-ŋ́*, as in /kúg-ŋ/ → *kúŋ*

'tree (sp.)'. But in other noun stems ending in /g/, the /g/ does *not* delete, but a vowel is inserted, as in /kúg-ŋ/ → *kúgúŋ* 'cooking place'. The stem **g** shows up in the plural forms for both below.

	class	citation form	plural	gloss
a.	Noun Class 3	kú-ŋ	kúg-úsí	'tree (sp.)'
b.	Noun Class 1	kúg-úŋ	kúg-è	'cooking place'

The environments for the /g/ in the citation forms of these two words are virtually identical, differing only in noun class. Whether the rule applies or not seems to be arbitrarily a function of what noun class the word belongs to. The pattern here can be traced back to the historical sources of these nouns. Comparing Kɔnni to its nearest relative Buli, we find that Noun Class 3 words of 'Proto-Buli-Kɔnni' did not have a **g** in the root, but a **k**. This **k** was originally a suffix indicating Class 3, but has been re-interpreted as part of the root. So the protoform *kʊ**k** became kúŋ, and the protoform *kʊk-sɪ became kúg-úsí. But in Noun Class 1, the protoroots DID have **g**. The form *kug-i became kúg-úŋ, and *kug-a became kúg-è. We can make sense out of this apparently arbitrary pattern by a historical approach.

Sometimes historical linguistics can even solve serious sociolinguistics problems. David Thomas tells of the Chrau orthography in Vietnam which first followed what seemed to be a fairly central dialect but some people from another very large area objected. Eventually, all agreed to follow the 'early Chrau' forms which the Thomases had reconstructed (and which were fairly close to what they had been using), and this satisfied everyone.

In this issue of *Notes on Linguistics*, Joe Grimes gives an overview of the value of historical linguistics in fieldwork, and cites a number of SIL-authored historical publications. Paulette Hopple's paper summarizes the contribution of comparative studies in Southeast Asia. Bob Longacre reviews SIL's contribution to Proto-Otomanguean studies, and looks at some current issues. I think you will find these to be an interesting set of articles.

References

Cahill, Michael. 2000/2001. Noun classes and phonology in Kɔnni. Journal of West African Languages XVIII.1: 49–69.

Hock, Hans Henrich and Brian D. Joseph. 1996. Language history, language change, and language relationship: An introduction to historical and comparative linguistics. New York: Mouton de Gruyter.

—Michael Cahill
International Linguistics Coordinator

■

The Value of Comparative Linguistics

Joseph E. Grimes
Cornell University and Summer Institute of Linguistics

Not long ago, in a country not all that far away, the sixth in a series of language survey teams was sent into an area. The first five had either looked at only part of the area, or had not been able to form conclusions on the basis of the limited kinds of data they had been directed to collect. This was not surprising; other linguists who had looked over the area had found their own data problematic too.

Fortunately, the surveyors for the sixth round went beyond what they were instructed to do. They looked at some of the sound changes, which showed a regular pattern. For example, two of the varieties examined show a neutralization of voiced obstruents to voiceless in the final syllable of noun stems and the final consonant of verb stems. Three other varieties retain root-medial /k/ between vowels, while the two first mentioned along with four others drop the /k/ and coalesce-like vowels that come into juxtaposition as a result of the loss into a single vowel with a falling tone. They also looked at noun classifier regularities and forms of the negative. That type of structurally based observation could have been a sound basis for a definitive interpretation of the survey data had they taken it further.

Systematic comparisons such as the ones they did, based on proven principles, accomplish one of two things: either they give a highly reliable picture of language relationships that is not scrambled by random factors like linguistic borrowing, or they confirm that language relationships in an area are indeed crisscrossing and snarled—but then they pinpoint what the areas of uncertainty are.

What seems to be forgotten when survey teams encounter such complex situations is that comparative linguistics is the most highly developed means we linguists have for getting our bearings in the many-sided world of language varieties. Typology and the theory of universals also orient us to thinking across languages, but in other ways; it is comparative linguistics that gives us our best handle on language diversity.

The heart of comparative linguistics is the identification of systematic, patterned similarities and differences across speech varieties—phonology, morphophonemics, morphology, syntax, and semantics.[1]

[1] Fascinating examples of how all these intertwine can be found in Buck 1949.

These regularities fit a model in which changes in speech behavior propagate over time to some but not all of the speakers of a speech variety. Who is or is not affected by the changes depends on factors like geographic proximity and social group identification, so that migrations and political fallings out or alliances play a role in the process, even though what results is best described by linguistics.

Changes appear all the time. Some may be accepted widely in a short time; others creep slowly through social space. Others are ignored, or have only local influence.

The net impact of all changes over time is that some speech varieties are very similar to each other, and others less similar. Still others do not show any patterned similarities at all, evidence either that they come from different streams of development or else that they parted company so long ago that the patterns have faded beyond our reach.

The relationships implied by recognizable regularities yield something close to the kind of family tree seen in biological studies.[2] The tree metaphor is often looked down on because it requires considerable commentary about anomalies in relationships that are not quite tree-like.[3] Still, trees are a reasonable idealization for most of what happens in linguistic change. Clusters in multidimensional space would be a better idealization, but linguists in general are not yet comfortable with that mode of description.[4]

We all wish we could reduce the complexity of factors that result in relative closeness or distance among varieties to simple numbers, as if linguistic relationships were a linear measure like distance on a map.[5] Comparative

[2] The family trees drawn by genealogists and animal breeders trace the ancestry of particular individuals through sexual unions that produce offspring. Those used in biological sciences such as virology trace populations (such as new strains of the HIV virus), not individuals. Language family trees are of the latter variety, showing junctures where speech communities diverge, not the speech behavior of individuals.

[3] A good example is Chafe and Foster's observation (1981) on the development of the Iroquoian languages, that Tuscarora and Cayuga diverged from the rest in the earliest recoverable split, but later, still carrying the effects of that change, Cayuga joined back with the rest in behaving like languages such as Seneca with respect to some much later changes.

[4] One possible line of thought is opened up in Grimes and Agard 1959 (the beginning of what Howard McKaughan later dubbed "phonostatistics"), footnote 12, where the results of comparative phonology are presented as clusterings of multidimensional vectors. Various means of interpreting such data are discussed in Gauch 1982 and more recent software packages for the display of numerical data.

[5] A linear measure is one like measures of length, where a centimeter on one part of the scale is the same as a centimeter on any other part of the scale. Much of the universe is nonlinear;

linguistics helps us come back to reality by laying out what the actual dissimilarities of behavior are and inviting us to examine the influences each has on communication.

Manifestations of dissimilarity also come up when we investigate intelligibility: the overall ability, unenhanced by special second language learning, of speakers of one variety to communicate with speakers of some similar varieties in spite of noticeable differences in their speech. Intelligibility is one of the areas where linguistics has a directly practical side: information about intelligibility has been used effectively in decisions about educational delivery systems, disbursement of benefits, assignment of interpreters to emergency services, and translation of documents from other languages. Its potential usefulness may be even wider. Intelligibility is implied directly in one of the key questions for linguistic surveys (Grimes 1974): for the area of interest, what optimal set or sets of speech varieties allow adequate communication with everybody in the area?

When speakers of two related varieties fail to communicate, there are reasons why. When they communicate successfully, there are also reasons why. The late Frederick B. Agard (1975, 1984) pointed out that many of those reasons can be traced directly to the way the varieties have developed over time. Agard even presented a typology of kinds of phonological change (the most straightforward kind of change to trace via the comparative method) in terms of what changes do to structural congruity among speech varieties—a major factor in intelligibility.[6]

Another way to think about comparative linguistics is as a means of highlighting the regularities that stand behind the diversification of languages. It is the best tool we have for filtering out the vicissitudes of language development from the underlying main trajectories, for recognizing the borrowings and analogies and taboo effects that confuse the picture as

---

that is, a unit on one part of a scale has a more complex relation to a unit with the same name on other parts of the same scale. Moving a light switch two millimeters at either of its rest positions has no effect on the switch's behavior; moving the same switch two millimeters just at the center of its travel makes the light go on if it is off, and off if it is on. Light switches and braking automobiles and falling space debris, computer viruses and snarling dogs, all exhibit nonlinear behavior.

[6] Agard himself was looking for possible structural ways to distinguish between languages and dialects, independently of intelligibility considerations. But as Milliken (1988) shows, Agard's typology of sound change types is a good predictor of intelligibility and lack of intelligibility, and thus sheds light on the specifics of why people do or don't understand each other.

over against the regular sound changes that define branches of a language family.

The picture of this that helps me the most is that of the Gulf Stream, a major network of offshore currents in the western Atlantic Ocean. Water warmed in the tropical Atlantic and the shallower Gulf of Mexico makes its way north parallel to the east coast of the United States, then veers northeast off Canada, and finally east to Norway and Britain, and south to the Canary Islands. Vast quantities of water follow this route; but if weekend sailors unaware of what is going on were to look just at the windblown ripples on the surface or follow the track of the rubbish floating there, they would get a totally erroneous picture of what is going on farther down. Borrowings and analogies and taboo effects are the flotsam and jetsam of language history; to see what is really going on we have to look deep.

I think there is a perfectly understandable reason why some SIL field workers shy away from comparative studies. They have heard that comparative work is arduous, boring, tricky to handle, of interest only to antiquarians, and that it takes time away from descriptive pursuits. The supposed time factor has perhaps been the main reason why few SIL-run language surveys support their conclusions with comparative results.

As a matter of fact, the well-known 80/20 principle applies to language comparison as much as it does to anything else. Eighty percent of the comparative picture can be laid out in 20 percent or less of the time it would take to do a full scale analysis. Put another way, it is the least revealing 20 percent of the comparison that takes up 80 percent or more of the linguist's time. For a useful initial read on an area, the time consuming 20 percent can be put off until someone comes along with the curiosity and persistence and know-how to tackle it.[7] It need not take long at all to plot the major changes that define the shape of language diversity, especially if there is one person on the team who has a good grasp of the developments that have already been worked out for the wider region.

Partial automation is another way to diminish the time needed while enhancing the accuracy. The quill pen methods that were developed to a high degree during the eighteenth to the twentieth centuries are still valid and useful, but there is so much to keep track of that scribal errors tend to

[7] It is interesting to note that Szemerényi's 1996 treatise on Indo-European devotes only a tenth of its space to evidence for the major segmental regularities, and about the same amount to morphophonemic and accentual regularities. Another tenth is a general introduction to Indo-European and comparative linguistics. Much of the rest of the book is taken up with reason after reason why some things don't fit the strongly confirmed general picture.

propagate themselves all through the analysis, and just finding things requires a lot of time.

Because it's high time somebody did something about it, I am working on an application for laptop computers that stores data from large numbers of speech varieties, then gets the linguist's informed judgments about how each datum is to be compared with others like it. It processes and stores the implications of those informed judgments in a form that is easy for linguists to navigate, permits alternative analyses to be explored and results to be revised, and presents the data in a form suitable as an appendix to a comparative monograph—or to a survey report that contains carefully constructed arguments about the relationships among the varieties surveyed, not crude guesses.[8] This will give an independent reading to compare with the results of intelligibility testing. The full comparative details are there also, waiting for whoever eventually tackles the nasty 20 percent.

As a matter of fact, a number of SIL linguists have taken the trouble to learn the comparativist's craft, and have made contributions to historical and comparative linguistics. It has not impaired their ability to also contribute in other areas. Among those I have worked with personally one could mention Pike 1950, the group that mapped out the Otomanguean family of Mexico to a time depth comparable to that of Indoeuropean (Longacre 1957, Gudschinsky 1959, Rensch 1976, and Bartholemew 1989); also Wonderly 1949, Turner 1969, Waterhouse 1969, Orr and Longacre 1968, Payne 1991, Savage 1986, and Mead 1998; in the references I give a scattered selection of others.

People undertaking graduate studies might do well to consider investing up to a sixth of their class time in understanding the historical dimension of language. It is the arena where the basic concepts of modern linguistics were first worked out in detail before a recognizable descriptive or theoretical linguistics emerged from it, so it is worth looking into if only for breadth. Besides, a few may get hooked on it, to the benefit of all of us.

REFERENCES

Agard, Frederick B. 1975. Toward a taxonomy of language split (Part 1: Phonology). Leuvense Bijdragen 64:293–312.

———. 1984. A course in Romance linguistics. Volume 2: A diachronic view. Washington DC: Georgetown University Press, especially pp. 41–47.

[8] The application is also designed to allow collaborative research to be conducted over the Internet by teams of scholars located in different places, as well as done on a single computer by a single individual.

Aschmann, Richard P. 1993. Proto Witotoan (Summer Institute of Linguistics and the University of Texas at Arlington Publications in Linguistics, 114). Dallas: Summer Institute of Linguistics and the University of Texas at Arlington.

Bartholemew, Doris A. 1989. The Proto Otopamean vowel system and the development of Matlatzinca. Mary Ritchie Key and Henry M. Hoenigswald, eds., General and Amerindian ethnolinguistics: In remembrance of Stanley Newman. Berlin: Mouton de Gruyter, pp. 345–363.

Bee, Darlene. 1965. Comparative and historical problems in East New Guinea highland languages. Linguistic Circle of Canberra Publications A 6.1–37. Republished in Howard McKaughan, ed. 1973. The languages of the Eastern family of the East New Guines Highland stock. Seattle: University of Washington Press, pp. 739–768.

Buck, Carl Darling, et al. 1949. A dictionary of selected synonyms in the principal Indo-European languages: A contribution to the history of ideas. Reissued by University of Chicago Press, 1988.

Bugenhagen, Robert D. 1993. The semantics of irrealis in Austronesian languages of Papua New Guinea: a cross-linguistic study. Ger P. Reesink, ed. Topics in descriptive Austronesian linguistics (Semaian, 11). Leiden, Vakgroep Talen en Culturen van Zuidoost-Azië en Oceanië, Rijksuniversiteit te Leiden, pp. 1–39.

Campbell, Lyle and David Oltrogge. 1980. Proto-Tol (Jicaque), International Journal of American Linguistics 46:3.205–223.

Casali, Roderic F. 1995. On the reduction of vowel systems in Volta-Congo. African Languages and Cultures 8:2.109–121.

Chafe, Wallace L. and Michael K. Foster. 1981. Prehistoric divergences and recontacts between Cayuga, Seneca, and the other Northern Iroquoian languages. International Journal of American Linguistics 47:2.121–142.

Clouse, Duane A. 1997. Toward a reconstruction and reclassification of the Lakes Plain languages of Irian Jaya. Karl J. Franklin, ed. Pacific Linguistics (Papers in Papuan Linguistics, 2) A 85:133–236.

Davis, Irvine. 1966. Comparative Jê phonology. Estudos Lingüísticos 1:2.10–25.

———. 1968. Some Macro-Jê relationships. International Journal of American Linguistics 34:1.42–47.

Gallman, Andrew F. 1997. Proto East Mindanao and its internal relationships. Linguistic Society of the Philippines (Special Monograph issue, 44).

Gauch, Hugh G., Jr. 1982. Multivariate analysis in community ecology. Cambridge: Cambridge University Press.

Grimes, Joseph E. 1974. Dialects as optimal communication networks. Language 50:260–269.

———, and Frederick B. Agard. 1959. Linguistic divergence in Romance. Language 35:4.598–604.

Gudschinsky, Sarah. 1959. Proto-Popotecan: A comparative study of Popolocan and Mixtecan. Indiana University Publications in Anthropology and Linguistics, Memoir 15, supplement to the International Journal of American Linguistics 25.2. Bloomington: Indiana University.

Hedinger, Robert. 1987. The Manenguba languages (Bantu A.15, Mbo cluster) of Cameroon. London: School of Oriental and African Studies.

Jensen, Cheryl. 1999. Tupí-Guaraní. R. M. W. Dixon and Alexandra Y. Aikhenvald, eds. The Amazonian Languages (Cambridge Language Surveys). Cambridge: Cambridge University Press, pp. 125–163.

Lee, Ernest W. 1974. Southeast Asian areal features in Austronesian strata of the Chamic languages. Oceanic Linguistics 13:643–68.

Lithgow, David. 1992. Using grammatical data to determine language relationships in Fergusson and Normanby Island languages of the Papuan Tip cluster. David M. Snyder and John M. Clifton, eds. Papers from 6ICAL on Languages from Papua New Guinea (Language and Linguistics in Melanesia, 23). Ukarumpa: Linguistic Society of Papua New Guinea and the Society on Pidgins and Creoles in Melanesia, pp. 119–138.

Longacre, Robert E. 1957. Proto-Mixtecan (Indiana University Research Center in Anthropology, Folklore and Linguistics, Memoir 5, supplement to International Journal of American Linguistics 23:4). Bloomington: Indiana University.

Mead, David. 1998. Proto-Bungku-Tolaki: Reconstruction of its phonology and aspects of its morphosyntax. Rice University Ph.D. dissertation.

Milliken, Margaret E. 1988. Phonological divergence and intelligibility: A case study of English and Scots. Cornell University Ph.D. dissertation.

Orr, Carolyn and Robert E. Longacre. 1968. Proto-Quechumaran. Language 44:3.528–555.

Payne, David L. 1991. A classification of Maipuran (Arawakan) languages based on shared lexical retentions. Desmond C. Derbyshire and Geoffrey K. Pullum, eds. Handbook of Amazonian Languages. Berlin: Mouton de Gruyter, Vol. 3, 355–499.

Pike, Kenneth L. 1950. Axioms and procedures for reconstructions in comparative linguistics—an experimental syllabus. Mimeo.

Rensch, Calvin. 1976. Comparative Otomangean phonology. Indiana University Language Monographs 14. Bloomington: Indiana University Publications.

Savage, T. Dale. 1986. A reconstruction of Proto-Southern Mindanaon. Studies in Philippine Linguistics 6:2.181–223.

Smith, Kenneth D. 1972. A phonological reconstruction of Proto-North-Bahnaric. Language Data (Asian-Pacific Series, 2). Santa Ana, CA: Summer Institute of Linguistics.

Szemerényi, Oswald J. L. 1996. Introduction to Indo-European Linguistics (tr. from Einführung in die vergleichende Sprachwissenschaft, 4th edition, 1990. Darmstadt: Wissenschaftliche Buchgesellschaft). Oxford: Clarendon Press.

Thomas, David. 1990. On early Monic, Vietic, and Bahnaric relations. Mon-Khmer Studies 16–17:177–179.

Turner, Paul. 1969. Proto-Chontal phonemes. International Journal of American Linguistics 35:1.34.

Unseth, Peter. 1989. An initial comparison and reconstruction of case suffixes in Surmic languages. Journal of Ethiopian Studies 22:97–104.

van den Berg, René. 1991. Muna historical phonology. James N. Sneddon, ed. Studies in Sulawesi Linguistics. Part II (NUSA: Linguistics Studies of Indonesian and Other Languages in Indonesia, 33). Jakarta: Universitas Katolik Indonesia Atma Jaya, pp. 3–28.

Wares, Alan C. 1968. A comparative study of Yuman consonantism (Janua Linguarum, series practica, 57). The Hague: Mouton.

Waterhouse, Viola. 1969. Oaxaca Chontal in reference to proto-Chontal. International Journal of American Linguistics 35:3.231–233.

Watters, David E. 1975. The evolution of a Tibeto-Burman pronominal verb morphology: A case-study from Kham (Nepal). Linguistics of the Tibeto-Burman Area 2:45–79.

Wonderly, William L. 1949. Some Zoquean phonemic and morphophonemic correspondences. International Journal of American Linguistics 14:1.1–11.

Joseph E. Grimes, 84-664 Ala Mahiku Dr. #191, Waianae HI 96792.
E-mail: Joe_Grimes@sil.org

■

Historical Linguistics in Southeast Asian Language Programs

Paulette Hopple
SIL—Southeast Asia Group

Ancient Traditions. Most of the major languages of South and Southeast Asia have a written tradition that is longer than that of English. To SIL members more accustomed to thinking in terms of a language situation with no written tradition at all, it can be a bit daunting to contemplate serving where orientalists are already familiar with the vast historical panorama of Asian traditions and are accustomed to the rich religious, intellectual, historical, artistic, architectural, and political complexities of recorded ancient civilizations. In the last 25 years we have increasingly invested in areas of Asia where written traditions have extended more than 2,000 years.

Language and culture represent the 'weight and glory' of an ancient tradition. An outsider may have a tendency to ignore what has gone before or to minimize its value simply because it is unfamiliar and seems irrelevant. Modern western civilization tends to equate 'the past' with 'without current value,' and to say 'the past is gone, but the future is what we make of it now'. Because future-thinking, goal-minded linguists and language developers can too easily ignore historical data that is highly valued by the communities they serve, it is incumbent upon us to consider the implications of those traditions toward applied linguistics such as orthography development, language planning, literacy, and even language survey. If development follows the natural configuration of the linguistic landscape, it can flow in the traditional direction of natural historical change and social value and thereby utilize existing motivations rather than creating foreign, ill-suited versions that may seem both odd and 'up-hill'.

What follows are some thoughts about how historical linguistics intersects with and informs typical linguistic tasks that most SIL linguists, at least in Southeast Asia, would expect to accomplish.

Diglossia. Burmese modern writing styles can range between forms that reflect ancient 900-year-old patterns to forms that represent modern spoken slang. The latter are typically found in dialogical text, whereas the former are socially appropriate for high state, political or academic occasions. A modern speaker when learning to read Burmese is aware that the written form of his language contains words, particles, and syntactic arrangements that would be totally inappropriate for speaking (unless he were addressing a very stuffy audience). Switching between two forms or styles is natural for

the educated speaker. So then we may wonder, how different must the older written form of a language be in order to be considered a completely different language? What can be considered merely a range of linguistic style? The former is what Ferguson (1959) defined as diglossia, two languages within one socially recognized language. Changing a normal written Burmese text into a spoken language text can result in as much as 75 percent lexical change in the text. The fact that such changes are regular, systematic, and expected means that whatever the actual level of 'inherent intelligibility', one would have to say that illiterate speakers, especially illiterate second language speakers, would find literacy a far different exercise than what most western linguists are taught to expect . The history of the language is coded in the script consisting of pictures of the shapes of letters, but even more dramatically in the social forms of use of the literate language.

Phonology. Historically, the comparative method in linguistics has centered around phonology because phonology provides an avenue from which to approach historical divergence between dialects and to reconstruct the history of the language family. On a practical level, historical phonology provides clues to the probable adaptability of materials such as translations, literacy lessons, and oral media. These expectations are based upon the assumption that the closer two languages are phonologically, the closer they will also be in other areas of semantics and syntax . While that assumption is debatable, it is relevant to the work of SIL in attempting to provide both training materials for national linguists and end-user literature to newly literate communities.

Appreciation of historical phonological changes related to tone within a language or even across a whole linguistic area can aid a linguist in understanding often perplexing phenomena. Tonal languages in mainland Southeast Asia come in many varieties. Some languages have fairly clear pitch differences, often with three phonemic levels and two or more contours. Additionally tone systems tend to combine pitch with other features of the glottis, such as glottalization, voice phonation (creaky, breathy, modal), and features of the oral cavity . Of the five separate language families in Southeast Asia, all have developed tonal features to some degree.[1] Tone in this region is regarded not merely as a category of pitch, but rather a bundle of characteristics associated with the syllable.

[1] The five language families referred to here are Tibeto-Burman, Tai-Kadai, Mon-Khmer, Hmong-Mien, and Austronesian. An Austronesian language known to have developed tone is Eastern Cham.

Cross-linguistic comparison of tone languages has shed light on both the historical development of specific tones and on the nature of tonal systems. Such studies of tonogenesis can be strategic for understanding tonal categories in a specific language and how those categories correspond to tone categories in other related languages, as well as for identifying trends within the systems, and evaluating such applied linguistic aspects as native speaker reaction to splitting or merging a tonal category in a proposed orthography.

In Southeast Asia, since tone contributes one of the more significant aspects of the linguistic landscape, an extended discussion of tone here will explain other points later in this paper. Tones appear to have originated from similar processes across very different language families, namely, the depletion2 in daughter languages of protosyllable consonants, both initial and final. Compensatory contrast is realized as changes in the syllable nucleus . Some languages such as Sino-Tibetan are reconstructed with two prototones, but others such as Mon-Khmer have no tones in the protolanguages. French botanist and orientalist André Haudricourt (1954) first explained tonogenetic principles when reconstructing how Hanoi Vietnamese, a Mon-Khmer language, acquired six tones.3 The protolanguage had three types of syllable final segments: open or nasal, voiceless spirants (*-s or *-sh which reduced to *–h in pre-Vietnamese times), and stops (which reduced to *-? by pre-Vietnamese times). The protolanguage had a syllable initial contrast in voicing of stops. The natural pitch of syllables with initial voiceless consonants is inherently higher than syllables with initial voiced consonants. This pitch difference is not phonemic, but sub phonemic, a matter of mechanical phonetic forces.

By the sixth century, pre-Vietnamese word-final *–h and *-? had been eroded, leaving only the previously subphonemic falling or rising pitch on the preceding vowel as the remaining contrast. Thus the consonants 'surrounding' the vowel provided a matrix to generate two tonal qualities. First, syllable pitch levels are generated from initial consonants and second, contours—the direction of pitch movement—are generated from depleted final consonants. Figure 1 and Figure 2 schematize this process.

[2] *Depletion* refers to a state of being lessened in content, quality or value. During tonogenesis there is a gradual erosion or decay of the syllable structure resulting in altered phonetic detail and phonemic contrast centered on a different arrangement of features.

[3] When Asian linguists refer to a specific number of tones, they refer to whole configuration of tone units, not phonemic *levels* of tone. In this, they differ from Africanists and Americanists. [Editor]

Figure 1. Three syllable types with initial and final consonants in Proto-Vietnamese (adapted from Matisoff 1973) [4]

	Level pitch	Falling pitch	Rising pitch
Higher pitch	pa	pas > pah	paT > pa?
Lower pitch	ba	bas > bah	baT > ba?

Figure 2. Contour tonal replacement of three syllable types in sixth century Vietnamese

	Level tone	Falling tone	Rising tone
Higher pitch	pa	pà	pá
Lower pitch	ba	bà	bá

At the point in history of Figure 2, pitch direction was contrastive but pitch level (high or low) was non-distinctive because the contrast was still realized in the voicing contrast of initial consonants. By the twelfth century, the contrast between initial voiced and voiceless consonants merged, resulting in the phonemicization of two pitch levels together with the previously phonemic distinction of pitch direction. What was a three-tone system in the sixth century split into a six-tone system by the twelfth century. Figure 3 schematically represents these changes.

Figure 3. Six tones in twelfth century Vietnamese (in current orthography)[5]

	Level	Falling	Rising
High	pa	pa'	pá
Low	pà	pã	pạ

The fact that many languages in Southeast Asia have succumbed to the same basic pressures of syllabic decay, with resultant tonal systems, provides a strong basis for practical solutions. For instance, a linguist might be interested in extending the use of a practical orthography across dialects where differences in surface tones obstruct readability. Reconstructed relationships between tonal systems have aided in the situation of Khmu, a very large Mon-Khmer language that is spread over five countries as well as dispersed among various western countries. At one point it was clear that

[4] In Figure 1 the symbol /T/ represents abstractly any final stop. In Figure 2 the grave accent [à] indicates falling tone here, and acute accent [á] indicates rising tone.

[5] For those interested in comparing their data, the Vietnamese traditional tone names are *nang, hơ'i, sắc* for the high series, and *huyền, ngã, nang* for the low series.

there are three major dialects, two of which are tonal and one without tone but which retains the use of distinctive initial and final consonants. A proposed orthography used the consonants, finding that it was possible for readers to interpret the consonants as tones.

This solution may seem strange to some linguists from other parts of the world, but a consonant solution to tone is what is also found in the thirteenth century Thai script system of modern Thailand. The Thai script recapitulates historical sound shifts by writing consonants that have tonal significance.[6] Modern abstract consonant classes actually make eleventh century linguistic sense—class 1 (voiceless aspirates), class 2 (closed glottis—voiceless stops and pre-glottalized nasals), class 3 (voiced consonants). A modern velar stop will have at least three different symbols, once for each tone class.

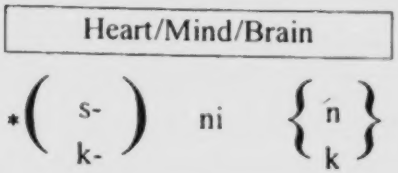
Semantic Reconstruction. Everybody knows that meanings of words change; however, the factors that cause these changes and the directions they take are complex and can appear unpredictable. By positing the presence of diachronic variation within a metaphor, however, we can begin to account for patterns in meaning change as well as ranges of associations and extensions of meaning. Synchronic variation in word meaning indicates a range of systematic ambiguity (with diachronic implications) for rich artistic effects in figures of speech, poetry, jokes, puns and so forth.

One particularly novel idea in Southeast Asia is that of a diachronic lexicon of a language or language family based on semantic domains constituting an inventory of central organizing metaphors. This lexicon would furnish insight and semantic breadth for such activities as translation, dialect adaptation or language standardization. Over the past twenty years, such a lexicon has been under construction at UC Berkeley's Sino-Tibetan Etymological Dictionary Project. Body part metaphors are one class of metaphors easily compared across languages since such data is regularly collected. The following example from Tibeto-Burman indicates how a body part metaphor is inductively reconstructed.

The common dichotomy in Indo-European languages between Brain (thought/perception) vs. Heart (emotion) is completely foreign to modern-day Tibeto-Burman somatic organization. One Proto-Tibeto-Burman reconstruction of a set of etyma for 'heart', 'mind', and 'brain' resulted in a combined protolexeme Heart/Mind/Brain as shown in Figure 4.

[6] The interested reader can find marvelous examples in Brown 1985 or more simply in Brown 1979:107–140 (Appendix II).

Figure 4. Proto-Tibeto-Burman— Heart/Mind/Brain (Matisoff 1978:212)



The macro protoform of Figure 4 is a reconstruction for a single etyma referring to 'heart/mind/brain' as its semantic scope in Tibeto-Burman. The data displayed in the languages in Figure 5 support this reconstruction and the observation that Tibeto-Burman languages have historically manifested this combined conceptual range of meaning.

Figure 5. Supporting data underlying the reconstruction in Figure 4[7]

Protoform	Language	Heart	Mind	Brain
*s- niŋ	Written Tibetan	snyiŋ	snyiŋ	
	Garo			təniŋ
	Nong			əniŋ
*s-nik	Written Burmese	<u>hnac</u> -lum		
*k-nik	Lisu	ni $_2$- ma$_3$		
*s-ni	Lahu	<u>ni</u>-ma		
	Mikir	ne		

The above chart demonstrates the basis for a single protoform with both a broad conceptual range and phonological variants. Such a study of Tibeto-Burman historical psycho-anatomy may have implications for translation of such concepts as 'flesh', the 'mind of the spirit', and so forth at a local language level, the language family level, and for areal studies. Since word meaning is organized around metaphors and we find those helpful to understand and predict semantic structure, so also can the historical study of semantic networks be valuable across related languages. Sounds and meanings change according to different principles, but like a married couple they usually are found living together.

Survey. It may be surprising to our language surveyors to learn that in Southeast Asia, the very STRUCTURE of a dialect survey word list used for

[7] Underlining indicates the reconstructed portion of a compound.

Tai-Kadai languages is framed upon protolanguage syllable patterns. This type of word list selection contrasts with semantically based word lists such as the Swadesh 100 or 200, or even to specialized regional /cultural lists.[8]

In Southeast Asia, the Tai-Kadai language family demonstrates widespread differences between phonology and lexicon from dialect to dialect and across the vast expanse of their geographic distribution, from Assam in India in the west, all across the Southeast Asia mainland, to Hainan island off the south coast of China. Yet despite those great differences, the family as a whole has maintained a stable, basic syllable structure that easily facilitates comparison.

Tone changes tend to mirror the historical divergence of Tai languages and provide a measurement of direction and degree of change, as well as principles of dialect grouping. William Gedney developed an historical tone matrix, now called the Gedney tone box following some of the basic assumptions of Haudricourt outlined above concerning Vietnamese (Gedney 1972). From the Proto-Tai[9] syllable structure of a) four sets of initial consonants and b) two types of syllables, it is possible to generate a matrix that explicates modern tones, as in Figure 6 below.

An earlier stage of non-tonal syllables has been posited by some scholars, similar to Proto-Vietnamese structures, but by the Proto-Tai period three Proto-Tai tones (labeled A, B, C) developed and are posited as one vector of the historical tone matrix of the open syllable.[10] The closed syllable, which is called in Tai the 'dead' syllable, is characterized by a further distinction of vowel length and its own type of tone (D1 or D2).

[8] One such wordlist is the 'Culturally Appropriate Lexicostatistical Model for South East Asia' (CALMSEA) proposed for Tibeto-Burman (Matisoff 1978:283-96).

[9] Tai is not the same as Thai. Tai refers to the broad historical branch of the language family called Tai-Kadai. The Tai branch of that family is composed of three main branches (Northern, Central, and Southwestern), which represent more of the commonly known languages (Thai, Lao, Zuang in China, etc.). The Southwestern branch of Tai is composed of the Tai languages found in Thailand, Laos, Burma, SW China, and some in N. Vietnam.

[10] Proto-Tai refers to the reconstructed level of all those languages within the classification of Tai. 'Proto-Tai Tone' refers to the reconstructed tone classes of Proto-Tai. Proto-Tai Tones A, B, C, D1 and D2 are tone classes found in modern-day languages. They represent tones in the proto-language, not just modern tones.

Figure 6. Proto-Tai tone box

	Proto-Tai Tone	Open Syllables A	B	C	Closed Syllables Short vowel D1	Long vowel D2
Initial consonants at time of tonal splits	Voiceless friction sounds *s, hm, ph, etc.	1	1	1	1	1
	Voiceless unaspirated stops, *_p_, etc.	2	2	2	2	2
	Glottal, *_ʔ_, _ʔb_, etc.	3	3	3	3	3
	Voiced, *_b, m, l, z_, etc.	4	4	4	4	4

The tones in Tai are typically referred to as A1, B4, etc., referring to their matrix position. This identifies the tonal category to the linguist without the bothersome details of the exact phonetic representation of that tone. It is a convenient way to reference the tonal systems of related languages using an abstract tonal classification. The numbers are succinct classifications of the Proto-Tai phoneme classes which have worked together to produce group effects on the tones. The pattern of tone class mergers can provide clues to language surveyors. Sometimes the group effects have not been unitary. That is, some of the phonemes in a class change in ways different from others, resulting in a tonal split within one abstract category, for instance a split in Tai Lue D1-4. This is shown in Figure 7, which demonstrates the tonal correspondences for two related Tai languages, Tai Lue and modern Standard Thai.

Figure 7. Two Tai language tone correspondences

Class of proto-initials	Initial consonant description	Proto-Tai tones				
Tone correspondences for Modern Thai						
		A	B	C	D1	D2
1	Voiceless Friction	MR		HF	LL	
2 & 3 (collapsed)	Voiceless Aspirated and Glottal	ML	LL			
4	Voiced		HF	HL		HF

Key	
HL	High level
HF	High falling
MR	Mid rising
ML	Mid level
LL	Low level

Tone correspondences for Tai Lue (data from Harris 1975)

		A	B	C	D1	D2
1	Voiceless Friction	MR				MR
2	Voiceless Aspirated	ML	LL	HF	LL	ML
3	Glottal					MR
4	Voiced	HL	ML	HL	HF / ML	HF

The degree of divergence and the distribution of tone patterns within the matrix provide a general tonal map to the language family as a whole. Broad patterns of tonal convergence are then used to classify historical branches within the language family.

Survey word lists use etyma belonging to each of the tone boxes. A sample list in Standard Thai is provided in Figure 8. Once one has the right matrix, and a modern test word whose syllable structure historically matches a particular box in the historical matrix, a measure can be made of the degree to which the modern daughter language diverged from the protoform.

BEST COPY AVAILABLE

Figure 8. Gedney tone box using Standard Thai (Harris 1975)

	A	B	C	D1	D2
1	huu 'ear'	khay 'egg'	khaw 'rice'	mat 'flea'	khaat 'broken, torn'
2	pii 'year'	paa 'forest'	paa 'aunt'	kop 'frog'	p□□t 'the lungs'
3	bin 'to fly'	baa 'shoulder'	baa 'crazy'	bet 'fish hook'	dɛɛt 'sunshine'
4	naa 'ricefield'	phii 'older sibling'	nam 'water'	nok 'bird'	miit 'knife'

Each tone box contains a battery of at least three test words that can provide an initial 'Rapid Appraisal' of dialect relationships. Thus we might add another survey tool to our dialect survey toolbox, one which has a firm basis in historical linguistic studies.

The data in Figure 8 is specifically for Standard Thai. Words in Standard Thai have already been reconstructed as etyma belonging to a particular cell. To explicate this a bit further, the word *huu* (rising tone) in Thai is derived from Proto-Tai syllable structures which have generated this modern tone. Using the survey trigger word *huu,* while eliciting in Thai for some other Southwestern Tai language, would presumably result in a cognate for 'ear', which has a tone that belongs historically to tonal category A1. By using at least three trigger words from each tone category (say ones that have derived from different Proto-Tai initial consonants within the same general class—*s, *hm, *ph) it should be possible to get a rapid understanding of how A1 tone category is realized in the target dialect being surveyed. Thus with only about 60 words elicited one can map out how the Proto-Tai tones have diverged in the language being surveyed.

The patterns between the boxes show the degree of divergence from Proto-Tai and the direction of historical changes. It is such regular patterns that historical linguists find most exciting and informative. Also the same linguistic/tonal patterns help us to understand the degree of language divergence, and likely intelligibility between languages, which is the goal of language survey.

Beyond Survey. The Gedney tone box analysis has proved useful in other kinds of Southeast Asian language situations. A study of Southern Thai[11] demonstrated tonal variants among thirteen dialects but all showed the same basic splits and mergers at very old levels. Interestingly, the tone system of Southern Thai was found to correspond as the probable source language for the modern system of writing tones of Standard Thai (Brown 1985). Standard Thai has a different tone system than the writing system would seem to indicate, resulting in the use of various rules to generate the modern correct tone from the written form. Such a rule is on the order of the English spelling rule '*i* after *e* except after *c*'. Knowing that there was a Tai language which served as the probable source language when the written form of Thai was developed provides an explanation for the modern misalignment of pronunciation with the written tones, and elevates the historical role of a section of the country and of a speech form that has been culturally sidelined with the development of the modern nation state.

Another benefit of the Gedney tone box has been in generating working hypotheses. For instance one student at a local university conducted a tonal dialect study of Northeastern Thai[12] for his thesis. He found that every district of every province had a different tonal map. For an SIL person who had an interest in the question 'What tonal pattern should the written form of Northeastern Thai handle?' these tonal configurations provide a linguistic base from which to generate working hypotheses about how districts might understand various test orthographies, hypotheses about what regions, provinces or districts may require special sets of lessons to use prospective reading materials, and hypotheses about the social history and migration patterns of people who endured forcible resettlement from Laos generations ago.

Summary. This article has painted in broad sweeping brushstrokes a view of the relevance of historical linguistic knowledge in Southeast Asia to both theoretical and practical linguistic enterprises—in phonology, semantic analysis, sociolinguistics, orthography, dialect survey, and translation.

[11] Southern Thai—6.5 million speakers, 13 dialects, no scripture, no written form of the present-day language.

[12] NE Thai in Thailand (also called Issan or Lao)—15 million speakers, some attempts at scripture translation, no standard written form in the Thai script, clearly a dialect extension of the same three major dialects of Lao inside Laos, one of which has a major written tradition but in a script not used in Thailand.

REFERENCES

Brown, J. Marvin. 1979. Reading and writing text. AUA language center Thai course. Bangkok: American University Alumni Association Language Center.

———. 1985. From ancient Thai to modern dialects. Bangkok: White Lotus.

Gedney, William J. 1972. A checklist for determining tones in Tai dialects. Studies in linguistics in honor of George L. Trager, ed. by M. Estellie Smith, 423–440. The Hague: Mouton.

Ferguson, Charles. 1959. Diglossia. Word 15.325–340.

Harris, Jimmy G. 1975. A comparative word list of three Tai Nüa dialects. Studies in Tai linguistics in honor of William J. Gedney, ed. by Jimmy G. Harris and James R. Chamberlain, 202–230. Bangkok: Central Institute of English Language, Office of State Universities.

Haudricourt, André. 1954. De l'origine des tons en viêtnamien. Journal Asiatique. 242.69–82.

Matisoff, James A. 1973. Tonogenesis in Southeast Asia. Consonant type and tone, ed. by Larry M. Hyman, 71–95. Los Angeles: Linguistics Program, University of Southern California.

———. 1978. Variational semantics in Tibeto-Burman: The 'organic' approach to linguistic comparison. Occasional papers of the Wolfenden Society on Tibeto-Burman linguistics 6. Philadelphia: Institute for the Study of Human Issues.

Paulette Hopple, 7500 W. Camp Wisdom Rd., Dallas, TX 75236.
E-mail: Paulette_Hopple@sil.org

■

BEST COPY AVAILABLE

Comparative-Historical Mesoamerican Reconstruction and SIL Personnel: Accomplishments and Problems.

Robert E. Longacre
SIL and Prof. Emeritus, University of Texas at Arlington

0. Introduction. This paper traces in summary form the role which SIL members have played in comparative-historical studies in Mesoamerican languages, reports a recent conference on that subject in Mexico City, and presents a few typical problems in the reconstruction of the Mixtecan language family as a branch of the broader grouping Otomanguean—which includes most of the tone languages of Mexico.

1. SIL studies in Proto-Otomanguean. Although there had been some comparative-historical studies in Mesoamerican languages (Swadesh 1947, Newman and Weitlaner (1950), my study Proto-Mixtecan (1957) was the first detailed study of this sort. It embraced three language complexes—Mixtec, Cuicatec, and Trique—and constructed a hypothetical forbear of the three: Proto-Mixtecan. Prior to this work there had been some confusion as to the position of Trique, which Jimenez Moreno in his linguistic map of 1937 had put into the Popolocan language family while classifying a fourth language, Amuzgo, as Mixtecan.

Sarah Gudschinsky, only two years later (1959) published a study, Proto-Popotecan, in which she reconstructed Proto-Popolocan on the basis of Mazatec, Popoloca, Ixcatec, and Chocho, then proceeded to relate this reconstruction to Proto-Mixtecan. The resulting reconstruction, which she dubbed Proto-Popotecan, was clearly a protem piece of work. It showed that Popolocan could be shown to be systematically related to Proto-Mixtecan, but there appears to be no particular historical layer especially involving only these two language families; Gudschinsky had simply related these two families as a first step towards what is now termed Proto-Omanguean as reconstructed later by Calvin Rensch (published 1976 after being in press for seven years).

Meanwhile, other branches of Otomanguean were being reconstructed: Proto-Chinantec by Rensch (1968), and Proto-Otopamean by Bartholomew (1965). Mexican scholars, especially Maria Teresa Fernandez de Miranda, were also pushing ahead with studies of other branches of Otomanguean: Fernandez de Miranda and Roberto Weitlaner (1961) in Proto-Chiapanec-Manguean (extinct languages on the southern edge of Mesoamerica). and

Fernandez de Miranda in Proto-Zapotecan. The latter work was delayed in publication due to the untimely death of its author, but her prepublication data were made available to Rensch, and her work itself finally saw publication in 1995. Some minor contributions were Proto-Mixtec, a study of Mixtec dialects by Cornelia Mak and Longacre (1960), and Proto-Chatino by William Upson and Longacre (1965).

2. Conference. At a recent symposium at the Autonomous National University of Mexico (August 28-September 2, 2001 under the auspices of the Institute for Anthropological Investigations) the topic was the genetic classification of Mexican Indian languages. There were sections on Mayan languages, on Zapotecan, on Hokan, and Mixtecan. Stephen Marlett of SIL challenged the very concept of a Hokan family of languages. He feels that the cognates are too few and the systemic reconstruction so scanty that there is little evidence for a Hokan language family.

At the same symposium, three SIL people were invited to make presentations regarding the Mixtecan language family: Barbara Hollenbach, Inga McKendry and myself. Hollenbach gave a truly masterful presentation of Mixtec tone in time and space, beginning with a sixteenth century catechism in Mixtec, and positing the town of Teposcolula, Oaxaca, as an early center where an attempt was made on the part of the Dominican friars to standardize Mixtec as a medium for teaching. She attempted to show that a verb prefix widely attested in colonial sources lost its segments and developed into a floating tone found in nearly all contemporary variants of Mixtec to mark present tense. She also demonstrated that an irregular tone sandhi pattern found in a cluster of towns developed via the movement of a high tone to the right. Contemporary Mixtec tone systems show wide variety especially in regard to systems of tone sandhi, and Hollenbach has attempted to show some systematic tendencies in this divergence.

McKendry read a paper on nasalization and palatalization in Mixtec with data from a wide range of Mixtec speaking villages. I read a paper concerning the main outlines of my 1957 reconstruction of Proto-Mixtecan and how various of us, especially Rensch, have modified it through the years since then.

3. Outstanding issues. In my reconstruction of Proto-Mixtecan (PMx) in 1957 as subsequently modified by Mak and Longacre (1960) the following consonants were postulated:

*t	*y	*k	*k^w	*w
*θ	*l	*x	*x^w	
*^{n}d		*^{n}g	*$^n g^w$	
*n		*m		

Rensch (1976) posited that on the common earlier layer, Proto-Otomanguean (POM), there was a palatalizing element *Y, which resulted in certain palatalized consonants in Proto-Popolocan. He also posited consonant clusters in POM as the source of the prenasalized consonants in PMx. He and I together felt that PMx is defined by two shared innovations relative to POM: (1) the merging of POM *Yt with *t and the merger of *Yθ (which he reconstructs as *s) with *Yθ and (2) the development of the prenasalized series in PMx. In brief, a series of palatalized consonants did not characterize PMx. With these defining innovations in PMx as a beginning argument it is possible to contend that Trique is part of the Mixtecan language family while Amuzgo is not. Trique shares in (1) and (2) above while Amuzgo does not.

But precisely here is where McKendry's studies are relevant. She argues, from her extensive study of palatalization in the various varieties of contemporary Mixtec that not all palatalizations can be limited to developments within particular Mixtec languages, but that some should be projected back onto the horizon of Proto-Mixtec itself. Note then the succession of mergers and splits in these developments:

(1) POM *Y plus consonant merged with the same consonant without the palatal element. Thus, PMx did not have a series of palatalized consonants.
(2) In those diverging PMx dialects which were to become Mixtec, a series of palatalized consonants emerged.
(3) Palatalization has continued to develop in the various Mixtec languages right down to the present time

In my publication of 1967 I trace various isoglosses in the diverging mass of PMx dialects. Some isoglosses include Mixtec and Cuicatec, others include Mixtec and Trique, while still others include Cuicatec and Trique. The latter are of special interest in that Cuicatec and Trique no longer are spoken in contiguous regions; they point to the time when the two populations were contiguous and shared a common life. On the other hand, Mixtec and Cuicatec, and Mixtec and Trique have apparently been contiguous throughout the history of the development of the three languages and their further internal divisions. In each case these isoglosses are not simply shared retentions from POM but are shared innovations on the horizon of the emergence of the three dialects of PMx which were to become three languages and finally three language complexes. Nothing was static at any time, then or now.

The dialect that was to become Trique was not a palatalizing dialect. The regular Chicahuaxtla Trique reflex of PMx **t* is *ch*, which for the most part displays complementary distribution, with a retroflexed allophone occurring before back and central vowels and a non-retroflexed allophone before front vowels. For obscure reasons, a phonemic contrast has developed in Chicahuaxtla Trique; four lexemes occur in which the retroflexed *ch* occurs before front vowels. But there seems to be no good reason to believe that the split of Chicahuaxtla Trique *ch* into two phonemes could be regarded as a 'palatalization'; in fact, the retroflexed form appears in four lexemes where it SHOULD be fronted by the front vowel but does not. Data from San Juan Copala Trique (SCJ) and San Martin Itunyoso Trique should be brought to bear on this problem. But apparently Trique can be characterized not as a palatalizing language but as a retroflexing (or 'anti-palatalizing'!) language.

The question as to what Cuicatec dialectology could contribute to this discussion will remain unanswered until such a study is made. At present Cuicatec data from a variety of villages where that language is spoken are not available. When I wrote Proto-Mixtecan I had word-lists made available to me by Marjorie Davis and since then a large Cuicatec dictionary compiled by Richard Anderson has been published (1983). Possibly a fresh beginning could be made with these materials. But this is still a far cry from the abundance of dialect data available for Mixtec and the availability of data from three Trique regions. Here as elsewhere comparative-historical reconstruction remains dependent on adequate synchronic studies.

Calvin Rensch (1976) in his revision of PMx as part of his overall POM reconstruction reduces the number of PMx vowels to four. This ties into the problem of the number of laryngeal phonemes that he posits in PMx. While I posited one laryngeal, Rensch has posited two such phonemes. With two laryngeals and postposed **m*, Rensch is able to reduce the vowel inventory for Proto-Mixtecan. It probably is a sure sign that Mesoamerican comparative linguistics has come of age that we find ourselves confronted with a laryngeal problem![*] In Chicahuaxtla Trique the two laryngeals *ʔ* and *h*, although in contrast on the contemporary scene, can by internal reconstruction be shown to once have been the same phoneme. Somewhat the same situation possibly holds for SJC Trique. But Hollenbach mentions that extra-short vowels are found in SJC Trique and that they occur in word final position and give the syllable a ballistic accent just as if it were closed by *ʔ* or *h*. She posits here an 'abstract laryngeal'. If Trique *ʔ* and *h* can be considered to be reflexes of one PMx laryngeal, it would be tempting to see in Hollenbach's 'abstract laryngeal' the reflex of Rensch's second PMx

[*] The question of laryngeals has been a long-standing point of contention in Indo-European reconstructions. [Editor]

laryngeal. At present, however, Hollenbach does not posit any such connection.

Any piece of historical-comparative reconstruction is an ongoing process, subject to revision by ongoing generations of scholars. If any reconstructed scheme remains unchallenged and unchanged for very long—whatever the defensive rear guard actions of older scholars such as myself—we do well to suspect that vigorous ongoing work has slacked off!

REFERENCES

Anderson, E. Richard and Hilario Concepción Roque. 1983. Diccionario Cuicateco. Mexico City: Instituto Linguístico de Verano.

Bartholomew, Doris. 1965. The reconstruction of Otopamean. University of Chicago dissertation.

Fernandez de Miranda, Maria Teresa. 1995. El Proto-Zapoteco. (Delayed and posthumous publication edited by Michael Piper and Doris A. Bartholomew) Mexico City: Instituto Nacional de Antropología e Historia.

———, and Robert Weitlaner. 1961. Sobre Algunas Relaciones de la familia Mangue. Anthropological Linguistics 3.71–99

Gudschinsky, Sarah. 1959. Proto-Popotecan: A comparative study of Popolocan and Mixtecan. Indiana University Publications in Anthropology and Linguistics, memoir 15. Supplement to the International Journal of American Linguistics 25.2. Bloomington, IN: Indiana University.

Longacre, Robert E. 1957. Proto-Mixtecan. Indiana University Research Center in Anthropology, Folklore and Linguistics, memoir 5. Supplement to International Journal of American Linguistics 23:4. Bloomington: Indiana University.

———. 1966. On Linguistic Affinities of Amuzgo. International Journal of American Linguistics 32:46–49

———. 1967. Systemic Comparison and Reconstruction of Middle American Indian Languages (accompanied by a linguistic map of Mesoamerica). Handbook of Middle American Indians, vol. 5: Linguistics, ed. by Norman A. McQuown, 117–159. Austin: The University of Texas Press.

———. 1968. Comparative Reconstruction of Indigenous Languages. Current Trends in Linguistics, vol.4: Ibero-American and Caribbean Linguistics, ed. by Thomas A. Sebeok, 320–360. The Hague: Mouton.

Mak, Cornelia and Robert E. Longacre. 1960. Proto-Mixtec Phonology International Journal of American Linguistics 26:23–40.

Newman, Stanley and Robert Weitlaner. 1950. 'Central Otomian I: Proto-Otomi Reconstructions' and 'Central Otomian II: Primitive Central Otomian Reconstructions'. International Journal of American Linguistics 16:1–19 and 73–81.

BEST COPY AVAILABLE

Rensch, Calvin. 1968. Proto-Chinantec Phonology. Papeles de la Chinantla, Série Científica. Mexico: Museo Nacional de Antropología.

———. 1976 Comparative Otomangean Phonology. Indiana University Language Monographs 14. Bloomington: Indiana University Publications.

Swadesh, Morris. 1947. 'The Phonemic Structure of Proto-Zapotec'. International Journal of American Linguistics 13:220–30.

Upson, William and Robert E. Longacre. 1965. 'Proto-Chatino Phonology'. International Journal of American Linguistics 31:312–322.

Robert E. Longacre, 7500 W. Camp Wisdom Rd., Dallas, TX 75236.
E-mail: Bob_Longacre@sil.org

■

REPORTS

28th LACUS Forum

Université du Québec à Montréal, July 31-Aug. 4, 2001

Shin Ja Hwang
Graduate Institute of Applied Linguistics

The 28th LACUS Forum was held on the campus of Université du Québec à Montréal, near downtown Montreal. LACUS conferences now have special themes for the presenters to focus on (The lexicon in 1999, Speaking and comprehending in 2000), and this year's theme: The nature of linguistic evidence. During the conference, there were three special one-hour plenary lectures, one presidential address at the banquet, one panel on evidence in linguistics, and over sixty papers, in twenty- and thirty-minute presentations.

Igor Mel'chuk gave the inaugural lecture on 'A formal language for morphology'. The other two guest lecturers were both from McGill University. Brendan Gillon spoke on 'Theory and evidence in linguistics: The nature of ambiguity, underspecification, deixis and vagueness', and Charles Boberg on 'Fact or opinion: A sociolinguistic view of native-speaker intuitions as evidence in linguistics'. I was particularly impressed by Boberg's presentation, a young professor who studied under Labov at U. Penn. and is still working with him on a book project. He gave suggestions for rapid and anonymous (R&A) survey methodology in collecting data from natural, unself-monitored speech, and warned us that native intuitions should always be compared with empirical data. Sheila Embleton of York University gave the presidential address on 'Bilingualism in contemporary Finland: Whither Swedish', which was very informative to me—not having much background in that area of the world. She showed how the use of Swedish, which used to be the prestigious language, is declining over the years, and gradually more and more people now speak English—practically a lingua franca in Europe.

The panel on evidence in linguistics (by Sydney Lamb, Carl Mills, Stephen Straight, and Victor Yngve) was rather brief—for one hour only—and not in-depth, but pointed out the neurocognitive basis of language that is the human brain as a type of evidence.

The rest of the conference papers ranged in topics from the kinds of evidence (3 sessions), discourse (2), morphosyntax (3), phonology, and semantic factors, to neurocognitive factors and iconicity—many in concurrent sessions. The participants ranged from the old timers and past presidents of

LACUS, e.g. Victor Yngve and Bill Sullivan, to the young or newcomers from as far away as Asia and Europe.

I will mention only some of the papers in this report. Given the conference theme on evidence in linguistics, a number of papers addressed that issue directly. Charles Ruhl (Old Dominion U), in his paper on monosemy, 'The evidence problem and some ramifications', listed four evidence types from Lamb's book, 'Pathways of the Mind':

1. The organs and processes of speech production
2. Texts, spoken and written
3. The processes of speaking, understanding, and learning
4. The neurocognitive basis of language—the human brain.

Lamb himself (Rice U.) focused on the use of neurological evidence, in addition to linguistic evidence, for a 'realistic approach to language'. Using a computer power point, also he did a very nice presentation touching on the use of quantitative tests.

Stefan Gries (Southern Denmark U.) compared three approaches to genitives in English to find legitimate evidence in linguistics empirically: (1) judgments by informed linguists, (2) judgments elicited from naïve native speakers, and (3) spoken corpus data. The results showed that (2) and (3) show much more similarity, both significantly diverging from (1). Again, we are given evidence that the use of native-speaker linguist intuition is not reliable. It was interesting, however, to find that some native-speaker intuition, i.e. of naïve speaker, can be useful in our studies. Stefan won the presidents' award for the best postdoctorate paper this year (which was first set up by Ken Pike)—apparently the second time for him after winning it last year. Now the LACUS board had to put a limit on how many times one can win the same award. Certainly his paper this year excelled, especially in directly addressing the conference theme.

The winner for the predoctorate award was Judith Yoel from Israel, whose paper was entitled 'Evidence for first language attrition of Russian sign language in Israel'. She illustrated how the sign languages are similar to spoken languages. Lexical items in the first language are gradually lost through the interference by those in the second language, and the degree of loss was based on sociological and psychological factors—just as spoken languages are. Vivien Ler Soon Lay from Singapore won the predoctorate commendation with her paper on the use discourse particle *lah* in Singapore English. The particle is frequently used in Singapore colloquial English and has a number of functions such as rapport, solidarity, persuasion ('Come

with us *lah*'). Using the Relevance Theory framework, she proposed it as a code marker of implicature.

Sarah Tsiang (University of Illinois) compared different verb forms like finite verbs and absolutive (participial) forms in Sanskrit data in 'Assessing evidence for discourse-motivated grammatical choice'.

Lilly Chen (Rice University) compared color symbolism between Chinese and English. Since 'red' is a positive color in Chinese culture, when the Dow is in the red, it means the market is up for the Chinese webpage readers while it is down for the U.S. readers. It can really be confusing! The stock market can be in dismal 'green' for Chinese, but it soars in the 'red'. Red symbolizes good-luck and happiness, as in 'red-event' (a wedding), but it also signals 'blood' and is associated with killing, danger, warning, red light, and bad. Probably the latter symbolism associated with blood is more or less universal across cultures. But we can see that the use of color terms in cross-cultural situation can be difficult. Proper names are often color terms in Chinese characters. My maiden name Joo is one of the characters for 'red', while Hwang is for 'yellow'. So I jokingly said to Bob Longacre that I must be 'orange', a mixture of the two colors.

Bob and Gwen Longacre were the only other participants from SIL beside myself. In his paper 'Proposal: A discourse-modular grammar of Biblical Hebrew', Bob argued for the study of verb forms according to discourse types since they often have different meanings and functions from one type to the next. He has been 'preaching' this idea for some time now, but not everyone in linguistics has been paying attention. This paper was yet another attempt for the same general goal, explaining two consecutive verb forms in the context of narrative and the 'P' complex (i.e. predictive, procedural, and instructional). He used a long handout with data including the Hebrew script, from the Book of Ruth to Jeremiah and Leviticus.

My paper (co-authored with Jonathan Lathers) was on 'Discourse structure of two parables', dealing with the parable of the workers in the vineyard (Matthew 20:1–16) and the wedding feast parable (Matthew 22:1–14). This was presented at GIAL Academic Forum, which helped me to sharpen my analysis. The original idea incorporating the 'points of reference' for Jesus' parables came from a paper that Jonathan did in my class in 1993. We planned to have a joint paper since then, but he has been too busy to work on the paper while preparing for and then being on the field. Since LACUS papers are severely limited in length—10 pages—we still would like to do a longer paper as a journal article, if Jonathan can spend time on the paper while he is on furlough this year in the U.S. Last week I submitted the written version of the paper to Ruth Brend, who is one of the editors of this

year's volume, and if the paper is selected through the review process it will be included in LACUS Forum XXVIII.

Studying parables in depth was very special for me—fun and rewarding. The two parables are of course very well known, and yet we seem to have difficulty in interpretation. The discourse analysis of the worker's parable helped me see the theme more clearly: 'God's mercy is equal for all, and we are all to be equally grateful'. The parable further points out that those who were first may be last since they are not as grateful as those who were originally last but are now first, being filled with gratitude for God's abundant grace. The whole parable (Matt.19:30–20:16) shows a clear chiastic, stepped-in, structure according to the content. The discourse structure of the wedding feast parable is more complex with an embedded discourse in the first part. It presents two themes: God's invitation for all of us to his kingdom, and the need for us to be changed and reborn (i.e. wearing special wedding garments). These two themes are directly related to the two points of reference in Jesus' audience—the chief priests and the Pharisees, and the multitude.

Next is a list of some of the papers that were very interesting to me:

- Samuel Navarro. An empathic perspective toward obviation in Cree
- Malcolm Finney. Universality of morphosyntactic processes: counterexamples from west African Kwa languages.
- David Bennett. Towards a better understanding of clitic systems.
- Pierre Larrivee. Meaning and grammatical functions: The case of locative sentence and verb modifiers in French.
- Joybrato Mukherjee. The scope of corpus evidence.
- Liang Tao. Pronouns and full NPs: Contextual dependency of reference accessibility.
- Toshiko Yamaguchi and Magnus Petursson. Anterior vs. resultative in Icelandic.
- Rafael Salaberry. Using lexical and grammatical aspect data to trace second language development.

In addition to presenting the paper on parables, I served as a session chair at the conference. Also I have served as a member of the executive board for the last three years and rated the abstracts. I will review papers in the areas of discourse and grammar in the selection process for this year's volume.

In closing, I would like to express my thanks to GIAL and SIL International (Academic Affairs) for funding me to participate in the 2001 LACUS forum.

Shin Ja Hwang, GIAL, 7500 W. Camp Wisdom Road, Dallas, TX 75236.
Email: shinja_hwang@sil.org

LACUS 2001: The Nature of Linguistic Evidence

Robert E. Longacre
SIL and Prof. Emeritus, University of Texas at Arlington

The Linguistic Association of the United States and Canada alternates yearly between conference sites in the USA and in Canada. This year the annual meeting was held in Canada at the Universite de Quebec a Montreal, July 31–August 4. The conference theme, as seen in the title of this report was 'The Nature of Linguistic Evidence'. Of the seventy-five some papers and addresses given, possibly half of the presentations could be considered to tie into the conference theme. Having attended several conferences wherein the conference planners hoped to secure discussion of a given theme—with often a majority of the papers ignoring it—I believe that the number of papers on evidence reflects considerable interest, perhaps even uneasiness, on this subject on the part of practicing linguists. Perhaps we are somewhat in the position of the centipede who walked along confidently enough until someone asked him which of his one hundred feet he moved first.

So this conference examined this old problem, especially the appeal to the intuition of the native speaker, but at the same time raised concerns from other quarters: neurolinguistics, corpus analysis, text analysis, and questions concerning whether linguistics can or should be a 'hard science'. In reporting the various papers which were presented, I will simply refer to handouts distributed during the meeting and to the meeting handbook, without the usual apparatus of bibliographical reference.

Victor Ingve is the continuing apostle of linguistics as a hard science, i.e., as opposed to soft science linguistics which is philosophy-based, employs non-physical concepts, has untestable theories and posits unobservable objects, and semiotic-grammatical foundations. He even questions the study of 'language' itself as an abstract entity. In contrast to all this, according to Ingve, Linguistics as a hard science studies people and sound waves that pass between them.

Ingve also states what he considers to be the ...

> ... four standard assumptions of (hard) science: (1) that there is a real world out there to be studied, (2) that it is coherent so we have chance of finding out something about it, (3) that we can reach valid conclusions by reasoning from valid premises, (4) that observed effects flow from immediate real-world causes.

It is interesting that in inveighing against philosophy, Ingve is driven to make some implicitly philosophical assumptions, as above. Is all this simply a variety of positivism? The emphasis on the real might indicate a type of realism. Be all this as it may. I do not disagree with his four assumptions, in that as a theist I believe in the kind of a universe that Ingve indicates. But the point is, his 'hard' science, like any such system of thought does not knock about in a philosophical vacuum!

Meanwhile, C. Hartnett's paper 'Morphological Patterns of Terms in Neuroscience and Quantum Physics' reminds us of the curious fact that physics, which we are accustomed to think of as a hard science, has grown quite philosophical and non-material. Comparing the vocabularies of quantum physics and the emerging discipline of neuroscience as two paths to reality, she finds that the former employs terminology that is somewhat more traditional than the latter. At this point the plot (linguistics on its way to become a science, or even a 'hard' science) perceptibly thickens. The great paradigm, physics as a hard science, has betrayed us. And if physics scarcely be saved where shall linguistics appear?

Douglas Coleman ('A Corpus Study in the (Non-) Physicality of Linguistic Observations') compares patterns of lexical usage in regard to the term 'data' in fourteen articles on Theoretical Linguistics (TL) and fourteen articles in Life Sciences (SL). Authors writing in the TL corpus 'tend to use "data" and "example" interchangeably' and are unlikely to identify precisely the observational source of their data—'very much less so than the LS authors'. Coleman's conclusion is that 'patterns of lexical usage identifying "data" as something physical (having objective existence) or not (existing subjectively only) will reveal significant differences between the TL and the LS corpora'. Again theoretical linguistics as currently practiced is seen to be less rooted in physical reality than are the life sciences. Of course, at this point, Ingve could rationally contend that theoretical linguistics is really the *bete noir* against which he inveighs and for which he wants to substitute linguistics as a hard science. But this is to condemn to the wastebasket some of the most prestigious current work!

Patrick Duffley ('Linguistics as an Empirical Science: the Status of Grammaticality Judgments in Linguistics') marshals arguments for discarding grammaticality judgments as highly unreliable linguistic data. His abstract ends with the words 'The conclusion will plead in favour of returning to the sounder methodological practice of confronting one's hypotheses with what is really there in attested usage'. I take it that his plea for a return to the examination of attested usage is meant to redirect our attention to text-based or at least corpora based studies.

For years now at LACUS meetings Sydney Lamb, and those of like persuasion with him, have stubbornly insisted on the relevance of neuroscience to linguistics. Behind communication, behind speech itself, behind conceptualization and even coherence itself lies the matter of what goes on in the brain. Neurolinguists such as Lamb believe that linguists should study 'not some illusory shared disembodied system, but the linguistic system of a typical individual. Moreover, that system must be seen as a neurocognitive system'. What Lamb hopes to achieve is a new realism in linguistics as seen in the title of his article 'Types of Evidence for a Realistic Approach to Language'. It seems to this reviewer that the payoff may be slow in coming. He speaks of immediate applications in 'Capacity testing' and 'processing speed' as means of testing models of neurolinguistic systems in particular individuals. He does not believe either that rules of grammar are innate or even included in the neurological base of language.

It is perhaps difficult for the neurolinguist not to be trapped in a kind of reductionism. Christian Wabl and Gernett Supp ('How Does it Come that a Lamb is called "lamb"?') skirt the edge of this pit:

> Linguistic units are neither symbols nor objects of any kind, they don't represent/don't stand for but refer to measured and compared percepts and at the same time they are part of the overall experience. Crucial terms under investigation have to be treated accordingly, have to be related to the appropriate neural phenomena like it is tried to be accomplished with consciousness, memory ...

Judging by the number of footnotes referring to articles on consciousness, the latter remains a major problem for the reductionist.

Such concerns as we have reported above have many ramifications. For example, the matter of corpus linguistics is discussed by J. Mukherjee ('The Scope of Corpus Linguistics') with the insistence that 'The line between corpus-based analysis of language use and intuition-based approaches to language system become blurred'. Still another variation on corpus-based analysis is to take as one's corpus televised interviews, as done by S. Schaefer in 'Prosodic Theory and Evidence in Oral Discourse'. Confronting spectrographic analysis with meaning in these oral materials, one can include this acoustic analysis as part of the enriched corpus. A skeptical note is sounded regarding the initial transcription of linguistic data by Rachel Selbach in 'Manufacturing Data: Transcription as the First Stage in Obtaining Linguistic Evidence'. She points out the fact that in a certain sense there is no 'raw' data. 'Transcribed texts of free/spontaneous speech' are, she insists, actually the first stage of analysis and are therefore not pre-theoretical or 'raw'. She reports how 'four different transcribers of different linguistic backgrounds produced consistently different transcripts of a non-

standardized highly variable Creole language'. She has a very thoughtful word for the linguistic realist:

> Science is quite removed from reality, it is one way of studying it. Linguistics is quite removed from language; it is our way of studying it. The study begins as soon as we dismantle the process of communication.

What of text analysis? As an enthusiastic proponent of textlinguistics, this reviewer has perhaps at times hawked it as the remedy for all our analytic ills as linguists. A moment's sober reflection will remind us that even here things are not quite that simple! While I continue to believe that the study of sentences in context is immensely superior to the study of sentences and their parts offered as 'evidence' apart from context, the textlinguist has his own set of problems. Some of these were highlighted by Sarah Tsiang in 'Ascertaining Discourse Motivations for Grammatical Choices'. Basing her remarks on Sanskrit texts Tsiang raised the question as to whether the choice of a given grammatical form over another is necessarily a significant choice in terms of discourse structure at all points in the development of a discourse? This reviewer heard her paper with mixed feelings. I reject 'free' variation in choice of grammatical forms, but believe that ultimately whether conscious choice or choice dictated by the conventions of discourse in a given genre, choices are on some level meaningful. But this article of faith is sometimes sorely tried in practice! And Tsiang let it all hang out.

Two other papers on discourse were presented, by Shin Ja Hwang and by myself. Hwang's presentation, 'Discourse Structure of Two Parables', was based on English translation of sections of Matthew's Gospel whose extant original is in Koine Greek. Recognizing that what counts as legitimate evidence in linguistics shakes down to intuition vs. use of available data, she opts for the latter as illustrated in her use of text from the ancient Greek of Matthew. She summarizes:

> The data-based approach is more fruitful especially when, as in this study, we analyze a structural unit of language that is larger than the sentence, or it may be the only approach available when we analyze a language with only written data but with no native speaker.

As a participant-observer in the sociological milieu that was LACUS 2001, this reviewer also read a paper that involved text analysis in an ancient language ('Proposal: A Discourse-Modular Grammar of Biblical Hebrew'). In it I unveiled a dream of writing a grammar of Biblical Hebrew that is squarely based on the study of verb forms in a language according to their usage in different types of discourses. All grammatical writing in Biblical Hebrew over the past thousand years has necessarily been corpus based and has consisted in citing forms from the Hebrew Bible as evidence for

grammatical claims and statements. But the forms have been cited without regard for discourse, i.e. contextual connections. I believe that a more solid sort of evidence can be offered by analyzing the textual corpus of a language as a system of discourse types in which, e.g., the aspect/tense/modal forms of the language each finds its proper setting. It remains to be seen whether a grammar so conceived and so dedicated can be shown to be more useful than its non-discourse-oriented predecessors.

This review of LACUS 2001 has been highly selective and has omitted mention of many good and provocative papers. I have narrowed my selection to papers especially bearing on 'evidence', and even here have omitted some bits and pieces of real worth. My own convictions regarding evidence in linguistics is based on the belief that linguistics occupies an ambiguous position between the physical and the social sciences. Since it deals with people and their vocal actions and interactions in society, linguistics is to this degree a social science. On the other hand, speech may be studied with considerable objectivity—even to the point of instrumental analysis—and in this respect linguistics resembles a physical science. Further resemblance to physical science is seen in the emerging discipline of neurolinguistics.

Robert E. Longacre, 7500 W. Camp Wisdom Rd., Dallas, TX 75236.
E-mail: Bob_Longacre@sil.org

■

REVIEWS

Possessives in English: An Exploration in Cognitive Grammar.
By JOHN R. TAYLOR. New York: Oxford University Press. 2001.
368 pp. Paperback $ 35.00.

Reviewed by CHAOQUN XIE,
Fujian Teachers University, China

Possessives in English (PIE) by John R. Taylor (JRT) is a significant contribution to the study of the English possessive and seems destined to become one of the keystone texts for theory and research in cognitive grammar. PIE is about a single morpheme in English and the constructions in which it occurs. It is argued that the cognitive approach to the English possessive is able to offer a more comprehensive and more insightful account of the data than theories associated with Chomskyan generative grammar.

Chapter 1 provides an answer to the question of why a whole book should be devoted to the English possessive morpheme that attaches to the ends of noun phrases to give forms like 'John's' and 'the city's'. JRT's answer is that in spite of their seemingly transparent nature, possessives present a number of special challenges for linguistic theory. Much of this chapter is devoted to preliminary discussion of possessive nominalizations, semantic issues and the language specificity of the English prenominal possessive, concluding that the distinctive character of the English prenominal possessive lies in the fact that English has grammaticalized this component of paradigmatic possessives.

Chapter 2 focuses on general points of contrast between cognitive grammar and generative approaches. Cognitivism, the language faculty, modularity, functionalism, mental processing, the scope of semantics, conventionality, lexicalism, and constructionism, all relate to learnability—an issue where contrasts between the approaches are particularly conspicuous.

Chapter 3 introduces some basic concepts of cognitive grammar. 'Symbolic unit' is a central one. According to Langacker (1987), the symbolic unit of cognitive grammar is the familiar 'linguistic sign', a notion that he extends both horizontally and vertically. Thus, not only are the words and morphemes of a language regarded as symbolic units; fixed idiomatic expressions also have symbolic unit status in what Langacker calls 'schemas'. JRT compares the views of generative and cognitive linguists

and concludes that a person's knowledge of a language contains a good deal of redundancy, with highly schematic knowledge coexisting with rather specific knowledge of how the general schemas may be instantiated.

Chapter 4 deals with the treatment of syntax and syntactic categories in cognitive grammar that bear on the possessive construction. There is a distinction, fundamental to cognitive grammar, between two kinds of linguistic unit: those that designate things, and those that designate relations. All linguistic units belong to one of these two categories. Nouns and nominal expressions designate things, while expressions of other syntactic categories (clausal, prepositional, adjectival, adverbial, etc.) designate relations. JRT sustains that the distinction is fundamental, not only to a characterization of syntactic categories, but also to the mechanism by which symbolic units combine to form units of increasing internal complexity.

Three distinctions are made between relational predicates: first, those that designate an atemporal relation versus those that designate a temporal relation, or process; second, those having to do with whether the landmark of a relation are things or are themselves relations; third, the possibility that the landmark may be 'incorporated' into the relational predicate. JRT then turns to an issue central to any theory of syntax, that is, the combination of linguistic units to form increasingly complex structures, in which the relations of modification and complementation are illustrated. As, JRT questions the traditional account for the role of determiners in which a noun phrase is headed by the noun, and articles are assimilated to the more general category of adjectival modifiers. He argues that the relational character of grounding would be captured by means of an unprofiled relation in the semantic structure of the determiner. The semantic structure of a determiner would be analogous to the semantic structure of an inherently relational noun like 'uncle'.

Chapter 5 discusses the constituent structure of prenominal possessives, focusing on some of the more controversial issues that arise within mainstream phrase structure and government and binding theories. JRT proposes a schematic structure for prenominal possessives, arguing that the import of the possessor nominal is to facilitate identification of the possessee, by mention of a reference point entity that is cognitively accessible, and from whose perspective the referent of the possessee nominal may be identified. It is in this chapter that the discussion of possessives really gets started.

Chapter 6 focuses on generative approaches towards prenominal possessives. The prenominal possessive construction in English has a number of distinctive properties of internal structure and usage range that set it apart from possessive constructions in many other languages. In spite of its

parochial character, the construction has figured quite prominently in the transformational-generative literature of the past two and a half decades or so. This chapter undertakes a critical review of some of these analyses.

Chapter 7 addresses three kinds of referential properties of prenominal possessives: definite-specific, indefinite-specific, and non-specific. JRT argues that prenominal possessives generally have specific reference and are nearly always compatible with definite reference; nevertheless, there are no grounds for excluding in principle the possibility that prenominal possessives may have indefinite, non-specific reference. Definiteness and specificity cannot therefore be regarded as inherent properties of the construction. In the two sections that follow, JRT examines attempts to derive the referential properties of possessives from general syntactic principles, pointing out that both accounts presuppose a conservative version of X-bar syntax, and both proceed on the factually inaccurate and problematic assumption that possessives invariably have definite reference. JRT argues that the failure of configurational accounts points to the need for an alternative approach, motivated by semantic and pragmatic considerations and that the referential properties of possessives fall out rather naturally from the reference account of the construction.

Chapters 8 and 9 are devoted to the reference point analysis in which the possessor nominal names a reference point entity that the speaker introduces as an aid for subsequent identification of the target entity denoted by the possessee. By considering the expected properties of reference point and target, JRT formulates a number of predictions concerning the possessor and the possessee. Confirmation of the predictions will tend to confirm the correctness of the reference point analysis.

JRT concludes that topicality offers itself as one factor behind the general impossibility of reversing the nominals in a possessive relation. This property is fully consistent with the asymmetry between the topical possessor and the non-topical possessee. Given the speaker's desire to uniquely identify the target, what determines the choice of reference point? In addressing the question, JRT draws attention to the restrictions on possessor nominals that are encapsulated in the Experience and Affectedness Constraints. Actually, what JRT strives to pursue is that, in addition to the requirement that it be topical, the possessor nominal needs to be such that it can provide reliable cues for the identification of the target.

Chapter 10, '*Ing*-nominalizations', addresses the use of the possessive morpheme in association with V-*ing* forms, as exemplified in three types of constructions:

Type A: the enemy's destroying of the city
Type B: the enemy's destroying the city
Type C: the enemy destroying the city

In terms of internal syntax, Type A is the most nominal of the *ing*-constructions. Types B and C preserve, in their internal make-up, a number of characteristics of the verbal expressions from which they derive.

Chapter 11 demonstrates that the distinctiveness of possessive compounds vis-à-vis non-possessive compounds turns out, in many cases, to be somewhat blurred and that the possessive morpheme in compounds has to be a completely different kind of entity from the possessive morpheme in prenominal possessives. What motivates the presence of the possessive morpheme in just some noun-noun compounds but not in others? JRT suggests that a possessive construal is favored just in case the compound exhibits features that are characteristic of prenominal possessives.

Chapter 12 deals with three other possessive constraints, each involving the use of a possessor phrase, namely, the prenominal possessive, the predicative possessive, and the postnominal possessive. In each case, the possessor phrase is used independently of a postposed possessee noun.

Chapter thirteen addresses the topic of possession. After clarifying the notion of possession, JRT presents several prototype accounts of possessive relations, concluding that these accounts share with taxonomic accounts a focus on the semantic relations between possessor and possessee to the neglect of the construction's discourse function. In the final section, JRT proposes that the reference point function involves a subjectivization of some aspects of paradigmatic possession.

Besides presenting a cognitive grammar analysis of the possessive, this book also undertakes a confrontation of the cognitive grammar approach with some alternatives—in particular those that have been pursued within the generative, and more recently, the government and binding paradigm, and its newest progeny, principles and parameters.

REFERENCE

Langacker, R. 1987. Foundations of cognitive grammar, vol. 1. Stanford: Stanford University Press.

Chaoqun Xie, Foreign Languages Institute, Fujian Teachers University, Fuzhou, 350007, Fujian, China. Email: xiecq1973@263.net

BEST COPY AVAILABLE

English transcription course. By M. LUISA GARCIA LECUMBERRI and JOHN A. MAIDMENT. London: Arnold; New York: Oxford. 2000. 156 pp. Paper $19.95.

Reviewed by KAREN STEFFEN CHUNG
National Taiwan University, Taipei.

English Transcription Course is a programmed series of lessons in representing spoken received pronunciation (RP) English in IPA symbols for both phonetics and English as a second language (ESL) students and teachers. What distinguishes this book is its emphasis on CONNECTED SPEECH PROCESSES, rather than just citation forms, which is all that many ESL learners are taught. The book offers two basic methods of giving the student practice: by providing passages written in (1) regular orthography for transcription into IPA; and (2) IPA transcription for reading aloud.

Each chapter introduces the student to a different type of allophonic variation that speakers tend to be little aware of, such as sandhi *r*. The student must learn to HEAR these processes and then to incorporate them into increasingly narrow transcriptions of the practice passages. The processes are described by the titles of the chapters— or 'Lessons'—3 through 8: 'Stress, rhythm and weak forms' (concentrates on reduced forms of function words, *and, there, can*); 'Sandhi *r*' (both regular linking *r*, *car of my own*, and intrusive *r*, *draw it*); 'Consonant syllabicity' (syllabic *n*, *listen*, and *l*, *little*); 'Elision' (alveolar plosive elision, *can't think*, and schwa elision, *history*); 'Assimilation' (*is she*, *red book*, *don't you*); and 'Glottaling' (replacement of *t* with a glottal stop, *not now*). Lesson One introduces 'Symbols and terminology', Two gives 'Transcription hints', and Nine offers 'Further practice'. There is also an introduction, an appendix with answers to all the exercises, a glossary, and a short bibliography of textbooks, workbooks, and pronouncing dictionaries.

Anyone who has assembled a book like this knows about the difficulty of finding suitable practice texts, mainly due to copyright restrictions. The authors solved this by writing all of their own material, a formidable task. They sidestepped a weakness of many practice books, i.e. insipid texts, by producing passages with interesting content—a necessity in maintaining student motivation. Although students can easily find texts on their own for transcription, and the authors encourage this, the key is getting help with transcriptions, which is where these self-produced texts with answer keys come in. They include personal anecdotes, descriptions of nature and places, and some surprisingly successful examples of creative writing, e.g. science fiction and household drama. The following is an excerpt from an anecdote

about a group of Italians on a train who were puzzling over the author's IPA transcription work (p. 51):

> ... the man said, 'It's him! He's doing it again! I wonder what that funny lettering is.' They all collected around me, peering over my shoulder. I couldn't resist the challenge. When I got off the train, I said in Italian, 'I hope you all have a pleasant day.' I wish I had had a camera to take a picture of the expressions on their faces.

One drawback is that the content is mostly narrative in nature, a story told by a single person, though there is one dialogue towards the end of the book. More examples of interactive speech, typical of what one would work with, for example, in discourse analysis, would be welcome.

The most difficult transcriptions to read with understanding (by this native speaker of U.S. English) are those in Lesson 7, 'Assimilation'. This lesson contains some very valuable information on, for example, assimilation of /t/ as in 'next summer' [ˈnɛks ˈsʌmə]. In some of these cases, the sound may in fact just be MASKED auditorily, even though the speaker may have their articulators in place for it; it might be useful to point this out in the text.

The absence of one allophonic process puzzles me—the pronunciation of /dj-/, /tj-/, /dr-/, and /tr-/ initials as in *due*, *tube*, *drink*, and *train* as [j̟], [ç], [dʒ], and [tʃ]. There is not a single mention of this in the book and no evidence of it in the transcriptions. Colloquial RP would sound rather odd without this allophonic process, and perhaps the authors may consider including it in a future edition. On the other hand, perhaps they felt the need to focus on a small number of processes and teach them thoroughly rather than overwhelm the reader. Still, it seems this particular one should be higher in the pecking order than, for example, glottaling, and that it could be squeezed into the chapter on assimilation.

Another area that could perhaps be addressed in a future edition is the linking of final consonants to initial vowels in the following syllable. This is something many foreigners (particularly in the Far East) are not taught. Instead they tend to begin every vowel-initial word or syllable with a glottal stop, making their speech sound choppy and halting. This kind of linking is also something native speakers are not necessarily very aware of.

A relatively minor point: perhaps || could be used as a symbol for a longer pause rather than |. The need for this was noticed when reading transcribed texts aloud—the biggest problems to understanding were ones of prosody and rhythm. The || would have cleared some of these problems up; e.g. [ˋaɪ ˈflʌŋkt ˈaʊt ɪn maɪ ˈfɜst ˈjɪə ˋ aɪ ˈdəʊn nəʊ waɪ ˋ] If the authors are

BEST COPY AVAILABLE

interested in a new, related project, perhaps they could consider producing a second volume that addresses intonation and prosody issues.

It is surprising that there apparently is no accompanying cassette tape or CD with this book. This would be of tremendous value particularly for instructors and students outside the UK, including those both in English-speaking and non-English-speaking countries.

This book will be useful and instructive to teachers and students learning to transcribe any dialect of English, though it is obviously most suited to those working in RP. The authors suggest (p. 3) two other books for those teaching standard U.S. English. Perhaps it would be useful to suggest books for other varieties of English, such as Canadian, South African, Australian, and even other UK dialects, if such books do indeed exist.

This review has in part ended up as a long list of suggestions. Perhaps it is because this book is basically so good and so useful to those involved in teaching and practicing phonetics that it is easy to get carried away with one's personal wish list. But this is a really well-thought out, well-designed, and meticulously produced volume. Phonetics and ESL instructors who are serious about cultivating in their students good pronunciation and clear, solid concepts of what that entails would do well to take a close look at this book.

Karen Steffen Chung, Mei Fu Building 7-501, Garden City, Hsintien, Taipei, Taiwan
Email: karchung@ccms.ntu.edu.tw

The Turkish language reform: A catastrophic success. By GEOFFREY LEWIS. New York: Oxford University Press. 2000. 190 pp. Cloth $60.00.

Reviewed by ALAN S. KAYE
California State University, Fullerton

Geoffrey Lewis, Professor Emeritus of Turkish at Oxford University since 1986, is renowned for his comprehensive grammar of the Turkish language (Lewis 1988). Very few Turcologists are as well-equipped as the author to examine the subject of the much-talked-about Turkish language reform of 1928. As is made clear in the introduction (p. 1), although there are many examples of the evolution of Turkish vocabulary, the book deals with much more than this, including the 'catastrophically successful' (cf. the subtitle of the tome) change from the Perso-Arabic script, which abounded in intricacies, to the Latin.[1] Lewis explains that the major aim of this Turkish

[1] The author calls it 'Arabo-Persian,' and states it is called that rather than 'Arabic' 'because it includes three letters *p*, *č*, and *ǰ*, that were added to the Arabic alphabet in order to represent the three Persian sounds not occurring in Arabic (p. 27, fn. 1). The author meant *g* for *ǰ*, since

language reform, closely associated with the name of the father of modern Turkey, Mustapha Kamâl Atatürk, was to eliminate Arabic and Persian grammatical features and loanwords which had been part of the language for centuries. The Turkish leader is quoted as saying (inside jacket): 'The Turkish nation, which is well able to protect its territory and its sublime independence, must also liberate its language from the yoke of foreign languages.'

The change of the Turkish alphabet took place in 1928 when the Perso-Arabic script was outlawed. This transformation is, I believe, the best example of linguistic engineering ever implemented. As the author notes, and we quite agree with him: 'Its intrinsic beauty aside, there is nothing to be said in favor of the Arabo-Persian alphabet as a medium for writing Turkish' (p. 27).[2] He goes on to point out that the four letters *?wlw* in Ottoman Turkish may be read in the following ways: *ulu* 'great' or 'possessors' (the latter being an Arabic loanword), *ölü* 'deed', *evli* 'married', *avlu* 'courtyard', and *avlı* 'stocked with game' (ibid.). The details given concerning the peculiarities of the Perso-Arabic script as used for Ottoman Turkish have been noted in Kaye (1996), to which I refer the reader for further details.

Not generally known and thus certainly worth repeating here, Lewis succinctly explains that spelling reforms in Turkey antedate Atatürk's action by many decades. The earliest attempt, in fact, was made in 1851 by one Ahmed Cevdet (p. 28). In 1862, the founder of the Ottoman Scientific Society, Antepli Münif Pasha, blamed the illiteracy problem in Turkey 'on the deficiencies of the alphabet' (ibid.). The following year, the Azerbaijani scholar, Feth-Ali Ahunzade, proposed the addition of vocalic graphemes, but the Ottoman Scientific Society dismissed the idea for various reasons. It took the charismatic leadership of Atatürk to pull off the actual feat of the linguistic change, and it is noted that he was preoccupied with a switch to the Latin alphabet during his residence in Syria between 1905 and 1907. He was already corresponding with a friend in Istanbul in Turkish before World War I using a French-based orthographic system (p. 31). The Turkish nation had to wait until August 9, 1928, however, for Atatürk to introduce the new Latin alphabet for Turkish at Gülhane Park. Two days after that, the following resolution was adopted (pp. 34–35):

Arabic has the latter but not the former (only a few colloquial dialects have a voiced velar stop). For the sake of completeness, one more letter should have been added, viz. *ž*. I believe, furthermore, that Perso-Arabic is far more common than the reverse designation, since the former term correctly implies that the Persians have modified the Arabic script, which they adapted to their own needs.

[2] See footnote 1

> To deliver the nation from ignorance, the only course open is to abandon the Arabic letters, which are not suited to the national language, and to accept the Turkish letters, based on the Latin. The alphabet proposed by the Commission is in truth the Turkish alphabet; that is definite ... The laws of grammar and spelling will evolve in step with the improvement and development of the language and with the national taste.

The rest of the story, as they say, is history. Lewis affirms that on November 1, 1928 the Grand National Assembly passed 'On the Adoption and Application of the New Turkish Letters' (p. 37). It should be kept in mind that it was decreed that on February 3, 1928, the Friday sermon in the mosque must be in Turkish (pp. 34-35). Judging by the official governmental literacy statistics, Lewis is right to call the reform a 'catastrophic success', since there was a 9 percent literacy rate in 1924, which increased to 65 percent in 1975, and 82.3 percent in 1995 (ibid.).

As the reader of this review might have already surmised, many details reported in this book revolve around Atatürk. There can be little doubt that 'he had a genius for synthesizing' (p. 43), and that he was widely read in the linguistics of his era.[3] It was surprising to learn, however, that he had a passion for unscientific etymology. One example of this indulgence discussed by the author is the derivation of *asker* 'soldier' from Turkish *asık* 'profit' + -er 'man' (ibid.). As Lewis points out, the word's etymology is ultimately from Latin *exercitus* 'army' (ibid.). Lewis reports that he even allegedly proposed etymologies for 'Niagara' and 'Amazon' (ibid.).

This is a fascinating work filled with important data and rich documentation. Lewis has done his homework well and the end product is a remarkable achievement. The book is extremely well researched and particularly well written. Only one error in an Arabic transcription was noted: the word for *guarded*, borrowed from Arabic into Ottoman Turkish, should be *maḥruus*, and not with a /ṣ /, which occurs three times as such (p. 7).

References

Kaye, Alan S. 1996. Adaptations of Arabic script. The world's writing systems, ed. by Peter T. Daniels and William Bright, 743–762. Oxford: Oxford University Press

Lewis, Geoffrey. 1988. Turkish grammar. Oxford: Oxford University Press.

Alan S. Kaye, California State University, Fullerton, CA 92634-9480.
Email: akaye@Exchange.Fullerton. edu

[3] Among the many books in his personal library were books by Otto Jespersen and H. W. and F. G. Fowler.

Learner-centredness as language education. By IAN TUDOR.
New York: Cambridge University Press. 1996. 279 pp.
Hardback $49.95, paper $19.95.

Reviewed by MARY MORGAN
SIL—International Administration

Learner-centredness as language education (LCLE) is one of the best books for language related linguistic fieldwork and adult education principles that I have come across. This book would have been a great help during the past four years that I myself was learning language as well as helping out new field workers in West Africa. It is highly recommended for all of our field libraries. It is a companion piece which reinforces the material on language learning in the LinguaLinks Library (SIL).

Two areas of interest to intercultural educators are discussed in LCLE. The first is the learner-centered approach to education and the second is language learning The book is a reference book that has questions for reflection and application that allow the reader to interact with the content.
The author gives a history of the learner-centered approach to education in the first chapter. Chapter 2 discusses learner training, chapter 3 looks at the objective needs analysis of the learner, and chapter 4 treats the subjective needs of the learner. The context is then discussed in which the learning takes place in chapter 5. Chapter 6 discusses self-assessment and fostering learning involvement. The final chapter looks at the teacher's perspective.
In the world of SIL, language learning is the basis for our fieldwork. The purpose of this book is to empower learners to understand the process of language learning, develop skills and strategies to work autonomously, and assess how they are doing. This purpose matches not only the needs of the field worker, but also the needs of the language learning consultants and language program coordinators in all of the countries in which SIL works.
In chapter 6 on self-assessment, section 6.5 (pp. 188-192) discusses developing strategic learning awareness. The conclusion for this section is the following:

1 Identifying the learning potential present in a given situation, activity or set of materials.
2 Selecting a limited number of learning targets
3 Knowing how to focus in on the target elements.

Chapter 7 looks at how to create in learners the ability to analyze their own language learning. Language learning activities such as exploring textual material, telling a story, using peer correction, correcting pronunciation, and written work are divided into three areas of analysis. The first has to do with

metacognitive strategies, the second the cognitive strategies, and the third the social/affective strategies. (These are adapted from the work of O'Malley and Chamot, *Learning Strategies in Second Language Acquisition*. Cambridge: Cambridge University Press, 1990.)

The author cites studies that are based on classroom and institutional situations, but the principles can be applied to the individuals who are on their own in a village situation.

Throughout the book, the learner-centered approach to education is carefully laid out. The principles and practices are given many examples. One can apply these to whatever kind of subject is under consideration. There are also case studies of traditional European, Asian, and Middle Eastern cultural expectations of teachers and students and the resistance that they bring to the learner-centered approach. These are helpful in acknowledging what can and cannot be changed in a learning situation.

The worth of this book is in the many suggestions for finding out what learners need as well as suggestions for strategies and activities for language learning. A second strength is the well-documented compilation of studies and materials. Further investigation is made easier by the references.

The author overstates and repeats throughout the book. This was a help to me because my concentration has been limited recently. For someone conversant with the material, this might be tedious. There are also many acronyms and I found myself looking back continually to remember what they stand for. The author chose for me the disconcerting convention of speaking of all teachers as feminine and all students as masculine.

I not only highly recommend this book for all of our field libraries, but I also recommend it for all of our language learning consultants. It would be useful also for teachers who are concerned about developing learners who can work autonomously as well as those who work within the classroom setting.

Mary Morgan, 7600 Glover Rd., Langley BC U2Y 1, Canada. E-mail: Mary_Morgan@sil.org

Tjurksko-Tatarskij etnogenez [Turko-Tatar ethnogenetic history]. By MIRFATYH Z. ZAKIEV. Kazan: Institute of Linguistics, Literature and Arts of the Russian Academy of Sciences. 1998. 624 pages.

Reviewed by YURI TAMBOVTSEV
University of Novosibirsk, Russia

It looks like the worst of times for the humanities, especially linguistics, in Russia are over. Some books have begun to be published. A fair example is the book by Mirfatyh Z. Zakiev. Though it is devoted to the Tatar language and ethnic history, it also touches on other Turkic languages (Bashkir, Chuvash, Karakalpak, etc.).

It will help the reader if some starting points are given for understanding M. Z. Zakiev's original and often controversial views on the development of the Tatar language and the ethnic development of the Tatars.

The Tatar people live on the Volga river in the center of Russia, forming their own national republic. At issue is where they came from and what their original homeland was. Consider first the classically accepted views on their origins. The name, Tatar, was already known in the 6th century BC among the nomadic tribes living near Lake Baikal in Eastern Russia (Siberia). In the 3rd century AD they began moving to the Volga river with the Huns and other nomadic tribes. Settling on the Volga, they mixed with the Finno-Ugric peoples. Later in the 7th to 8th century AD the Bulgarian tribes, who were also Turks, came to settle on the Volga river. In the 10th century AD all these tribes formed a united state called Volga Bulgaria, which later perished under the attacks by the Tatar-Mongolian hordes of Ghenghis Khan. In the 13–15th century AD these tribes were united under Mongolian rulers in one state under the Golden Horde. When that state collapsed, the Tatars divided into several separate states, giving rise to different Tatar dialects: Kazan, Astrahan, Crimean, and Siberian. The book under review mainly deals with the Kazan dialect of the Tatar language. The Siberian Tatar dialect is considered elsewhere (Tambovtsev, 1994a, 1994b).

This book by M. Zakiev is one of a series of books and articles dealing with the history of the Tatar language and the ethnic history of Tatars (c.f. Zakiev, 1995a; 1995b; 1995c; 1995d; 1996; 1997a; 1997b; 1998). In this review only the gist of Zakiev's basic ideas on Tatar ethnic genesis and development is given. One should try to read this book in Tatar; it is another step of democracy in Russia that such a huge book has come to be published in one of the languages of Russia. In reading the book, two things were surprising: 1) new and unexpected theories are presented, 2) Zakiev is widely known by his solid numerous books on Tatar syntax but this book is

not on syntax. Let's just mention the first and the last of his nine books on syntax (Zakiev et al., 1971; Zakiev, 1999). While every scholar fully accepts Zakiev's ideas on syntax, there is widespread opposition to his ethnogenetic views—for examples see also Otkrytoe 1996a; 1996b.

In the book under review M. Zakiev argues that the ethnic name, Tatar, originates from two old Turkic roots *tat* 'foreign, alien' and *er* 'people' (p. 17). Unlike other Tatar linguists, he claims that anthropological, ethnic, and cultural features show that the Tatar Mongolian hordes were not the forefathers of the Tatar people. He categorically stresses that Tatar ethnic roots are much older and akin to Bulgars (p. 21). It should be mentioned that many of Zakiev's followers (R. A. Jusupov, K. Z. Zinnatulina, Sh. N. Asylgaraev, F. S. Safiullina, F. Chalil, L. Shakirova, R. Nafigov, G. Ahundov, A. Ahmadullin, F. Urmanche, G. Ahunov, D. Salimova, A. Karimullin, A. Minozhetdinov, S. Alishev, etc.) support his revolutionary point of view against the classical axiom that there was a 'great peoples' migration' in the 4th century AD when nomadic Mongolian Huns gained control of a large part of central and eastern Europe, pushing the peoples of Mongolia, Kazahstan, and Asia into Europe. According to Zakiev's ideas the Turkic tribes lived in the eastern part of Europe long before they were led by the Huns.

Actually, the whole mass of Turkic tribes were called Huns. He claims that old Greek authors mention the Huns as the settlers of Europe in the second century AD but they never mention that the Huns came from Asia. Zakiev puts forward the idea that the Turkic tribes moved to the west and got to Europe as a result of their fight against Greek and Roman colonization. Zakiev's followers are sure that he is quite correct in squashing the myth of Old Turkic forming only in the first millennium BC when the common Turko-Mongolian branch of the Altaic language family divided. Zakiev even finds traces of the common Turkic language in the languages of the American Mayan Indians, who moved from Asia to America fifteen to twenty thousand years ago. He also finds Turkic features in the Shumerian language of the 6th century BC and criticizes those scholars who think that Old Turkic tribes were Mongols. Non-Mongol skulls found in the oldest graves of Central Asia and Altai argue against a Turkic identification. He does not believe that any European tribes lived in Old Time Asia (Urmanche et. al., 1998: 21–22) and attributes such skulls to Old Turkic people.

Zakiev and his followers strongly criticize the idea that several waves of settlers came to the area of the Middle Volga river and the southwest Urals: first Finno-Ugric tribes, then Iranic, then Baltic, then Ugric (proto-Hungarians), then Bulgaric, and finally Kypchak-Tatars. One of Zakiev's interesting theories is that the proto-Hungarian people never spoke an Ugric

language. He refers to ancient Arabic and Persian travellers who mention the people of Mazhgar speaking a Turkic language. Zakiev connects the term Mazhgar to the Tatar tribe of Mishar. Thus, according to him, the Mishar people were not part of the proto-Hungarians, a viewpoint confirmed by my own statistical methods, which analyze consonant and vowel patterns that show that they were closer to Turkish (3.90), Uzbek (4.37), Kirgiz (4.38), and Kazah (4.79). In terms of similarity, Mishar was not as close to the Baraba dialect of Siberian Tatars (5.12) as one might have expected. The lack of proximity may be explained by Zakiev's theory: if Kazan Tatars had their origin on the Volga river, and Baraba Tatars in Siberia, then surely one should speak of the Baraba dialect and the Kazan dialect as two languages rather than as two dialects of one language.

Actually, in terms of logarithmic distance as calculated by my statistical methods, Kazan Tatar is also closer to Hungarian (6.83) than to such Turkic languages as Ujgur (6.84), Hakas (7.49), Karakalpak (7.50), Azeri(7.64), and Altaj-Kizhi (10.20). At the same time Kazan Tatar is much closer to Hungarian (6.83) than to the other Finno-Ugric languages, e.g. Northern Mansi (15.59) or Konda Mansi (17.91) of the Urals. However, Kazan Tatar is closer to Hungarian than the Finno-Ugric languages of Volga that have been in language contact for many centuries: Mountain Mari (12.68), Lawn Mari (11.30), and Erzja Mordva (10.53). These numbers may lead us to suppose that the sound chain of Modern Tatar is not as similar to the Finno-Ugric languages of the Volga region as one might expect.

The calculated values also reveal a certain similarity of Hungarian to the Turkic languages that still awaits explicit explanation. Nevertheless, it is a fact that Hungarian is not seen to be close to the Ob-Ugrian languages when one considers the typology of the distribution of consonants. If it really were part of the Ob-Ugrian subgroup, together with Mansi and Hanty, one would expect greater similarity. In fact, one can see that it is not close to neither Northern Mansi (15.91) nor to Konda Mansi (16.32). Hungarian is also typologically distant from Northern (Kazym) Hanty (16.92). Therefore we can see that the modern state of languages cannot explain all the influences in the past, but can give certain clues to them.

Zakiev strongly rejects the so called Alano-Sarmatian hypothesis according to which the ancient tribes of Alans were of Indo-Iranian origin. He points out that the linguistic material was too limited since only the Indo-European languages were considered in making the Indo-Iranian claim. Analyzing ancient sources, Zakiev found that they were more similar to Turkic than to the Indo-European languages. Therefore, he believes Scythians and Sarmats to be Turkic peoples. This allows Zakiev to speak about Scythian-Sarmatian roots of the Tatars.

Zakiev remarks quite correctly that history does not answer the question of where the ancient peoples (Avars, Hazarians, Bulgars, Pechenegs, and Kypchaks) disappeared to. He rightly believes that peoples do not appear from nowhere and disappear to nowhere. Actually, this is the exact question I used to ask my teacher of history—who had no answer. Zakiev thinks that all these peoples continued to dwell in the territories where the Turkic people used to live, and still live now. He argues that if Indo-Iranian tribes had really lived in these territories, they would have left more traces and peoples, not just the small group of Osetian people.

There is one more argument of Zakiev: if Scythians spoke an Indo-Iranian language, why didn't the ancient historians note its closeness to the Persian language? He stresses that all Scythian words should be reconsidered on the basis of the etymology of Turkic roots. He gives his own rather convincing explanation for the origin of all known Scythian words on a Turkic basis. At the same time he does not believe Osetins came from the Scythians.

Zakiev argues that Indo-Iranians never lived in the territory of Eastern Europe. Considering the origin of Alans, he comes to the conclusion that they are of Turkic origin. He finds that the well-known Zelenchuk epitaphy makes more sense when it is analyzed from the point of view of Turkic languages than from the point of view of Indo-Iranian languages. My investigation of the typological closeness of the Ossetian language to the Turkic languages on the basis of its speech chain shows the following distances: Uzbek (5.64), Kazah (6.37), Turkish (6.65), and Azeri (6.75). Actually, only Persian is closer to Ossetian (4.83); the other Indo-Iranian languages are at the same distances as the Turkic languages, or even farther: Ossetian-Gypsy (6.07), Tadjik (6.22), Hindu (9.50), Sanskrit (10.32), Bengali (10.47), and Marathi (11.08), etc. The mean distance between Ossetian and Turkic, and Ossetian and Indo-Iranian languages can show which group Ossetian is closest to: the mean distance between Ossetian and the Turkic group (9.19) is greater than that between Ossetian and the Indo-Iranian languages (8.72)., Although this difference is not really significant, it may still mean that Zakiev's theory, though it sounds rather unusual and even strange, has some foundation. For comparison, let's note that the mean distance between Tatar (Kazan) and other the Turkic languages is much less (6.17), while the mean distance between Tatar (Kazan) and the Finno-Ugric languages is twice as large (11.92).

It has not been my undertaking to criticize or support Zakiev's theory. Like every theory it has weak and strong points. One must consider properly and carefully the relative strength of his points. My main interest has been to draw the attention of Western scholars to this new theory. It is to be highly recommended for discussions and assessment.

REFERENCES

Otkrytoe [Open letter]. 1996a. Otkrytoe pis'mo dejatelej nauki i kul'tury Tatarstana Prezidentu Respubliki Tatastan M. Shajmiebu [Open letter from representitives of science and culture of Tartarstan to the President of the Tartarstan Republic M. Shajmiebu]. Kris 17 May 1996.

———. 1996b. Otkrytoe pis'mo dejatelej nauki i kul'tury Tatarstana Prezidentu Respubliki Tatastan M. Shajmiebu [Open letter from representitives of science and culture of Tartarstan to tne President of the Tartarstan Republic M. Shajmiebu]. Altyn Urda 15(85) May 1996.

Tambovtsev, Yuri. 1994a. Dinamika funktsionirovanija fonem v zvukovyh tsepochkah jazykov razlichnogo stroja [Dynamics of phonemic functioning in the sound chains in languages of different structure]. Novosibirsk: Novosibirsk University Press.

———. 1994b. Tipologija uporjadochennosti zvukovyh tsepej v jazyke [Typology of orderliness of sound chains in language]. Novosibirsk: Novosibirsk University Press.

Urmanche, F. I., and Mahmutov H. Sh. 1998. Academic Mirfatyh Zakiev. Moskva: Insan.

Zakiev, Mirfatyh Z. 1995a. Tatary: Problemy istorii i jazyka [Problem of history and language]. Kazan: AN RT.

———. 1995b. Problemy etnogeneza tatarskogo naroda [Problem of the enthigenetic development of the Tatar people]. Materialy po istorii tatarskogo noroda, 12–94. Kazan: OGN ANT.

———. 1995c. Ob etnonime 'volzhskie tatary' i 'bulagaro tatary' [About the ethnic name 'Volga Tatars' and 'Bulgaro- Tatars']. Problemy lingvoetnoistorii tatarskogo naroda, 3–8. Kazan: OGN ANT.

———. 1995d. Etnogenez i osnovnye vehi razvitija bulgaro-tatar. Problemy lingvoetnoistorii tatarskogo naroda, 9–27. Kazan: OGN ANT.

———. 1996. Is there any need in Paleoturkology? Jazyki, duhovnaja kul'tura i istorija t'urkov: traditsii i sovremennost', 21–23. Trudy mezhdunarodnoj konferentsii v 3 htomah, 9–13 ijun'a 1992. Tom 1. Kazan: ANT.

———. 1997a. Tatarskij jazyk [the Tartar language]. Jazyki mira [Languages of peace], 357–372. Tjurkskie jazyki [Turkic languages]. Moskva: Indrik.

———. 1997b. Etnojazykovaja sushchnost [Ethnolinguist' alan. Etnogenez karachaevtsev i balkartsev, 12–22. Materialy 'Kruglogo stola' (sent'abr' 1994 g. Karachaevsk). Karachaevsk.

———. 1998. Toerki Tatar etnonijasena keresh. Voprosy funktsionirovanija, dialektologii i istorii tatarskogo jazyka, 9–25. Kazan: ANT.

BEST COPY AVAILABLE

———. 1999. Tatarskaja grammatika, vol.3: Sintaksis. [Tatar grammar, vol.3: Syntax]. Moskva: Insan.

———, and H. R. Kurbatov. 1971. Sovremennyj tatarskij literaturnyj jazyk: Sintaksis [The modern Tatar literary language: Syntax]. Moskva: Nauka.

———, and S. H. Alishev. 1995. Vvedenie [Introduction]. Materialy po istorii tatarskogo naroda, 3–11. Kazan: OGN ANT.

Prof. Yuri Tambovtsev, University of Novosibirsk, Dept. of English and Linguistics, P.O. Box 104, Novosibirsk-123, 630123, Russia. E-mail: yutamb@hotmail.com

A functional approach to linguistic change through language contact: The case of Spanish and Otomi. By DIK BAKKER and EWALD HEKKING. Working Papers in Functional Grammar No. 71, 1999. 31 pp.

Reviewed by DICK WATSON
SIL—Graduate Institute of Applied Linguistics

At a special linguistic forum on February 8, Professor Mufwene of the University of Chicago presented a paper questioning the use of the term 'Creole' and describing the spread of Bantu across Africa from the viewpoint of language in contact. Whereas classical linguistics, and Bantu linguistics in particular, has been concerned with the historical comparative approach from a genetic bias, Professor Mufwene asked why more is not done to describe language change from a contact point of view.

In fact, there is a growing literature on language in contact, although mostly considering sociolinguistic factors. Therefore I was impressed by this functional approach to linguistic change through language contact, exemplified by Spanish and Otomi. The authors demonstrate that borrowing reflects important facts about the structures of both languages that cannot be learned by studying either language alone.

Following a brief, understandable, and interesting description of Otomi of Mexico, the authors discuss not only social factors, but a variety of linguistic changes through borrowing and structural reasons for them. Both content words and function words are examined. Their thesis is also supported by showing that Quechua has followed a somewhat different pattern in its borrowing from Spanish. For example, the vigesimal numerals of Otomi have given way to the decimal numerals of Spanish, whereas the decimal numerals of Quechua have not.

We are hearing more about grammaticalization as a diachronic process, but 'degrammaticalization' can occur through language contact as exemplified in Otomi. It is suggested that 'in general, functional factors motivate the

changes while formal factors of the grammar may put constraints on the kind of changes that may take place' (p.27).

This article provides a practical model for anyone wanting to describe language contact, both as a means of learning more about the structures of the languages involved and as a means of documenting the current stage of language change in progress. It is also a good example of 'function' in Functional Grammar.

Dick Watson, 7500 W. Camp Wisdom Rd., Dallas, TX 75236. E-mail: Dick_Watson@sil.org

ADVANCED LINGUISTICS: ROLE AND REFERENCE GRAMMAR
2 April–30 April 2002 * Horsleys Green

This course will introduce participants to the theoretical framework, known as Role and Reference Grammar, which has been set out by Professor Robert Van Valin of State University New York. The course will take you through the theory from basic structure of the clause to the interface with semantics and pragmatics. It will be a taught course with readings and assignments.

Role and Reference Grammar (RRG) as a syntactic theory and functional approach to language analysis offers a significant alternative to standard theories of generative syntax which are structure based and require multiple levels of abstraction from the surface form. RRG is a monostratal theory and posits only one level of syntactic representation, the actual form of the sentence. RRG is rooted in the communicative function of language and assumes that syntactic structures have a semantic basis. In terms of the abstract paradigmatic and syntagmatic relations that define a structural system, RRG is concerned not only with relations of co-occurrence and combination in strictly formal terms but also with semantic and pragmatic co-occurrence and combinatory relations. Consequently, RRG is a theory of syntax that takes you very comfortably from the morpheme to the discourse structure of a given language. Contact ETP_UK@sil.org for more details.

From the RRG Website: RRG grew out of an attempt to answer two basic questions: (i) what would linguistic theory look like if it were based on the analysis of Lakhota, Tagalog and Dyirbal, rather than on the analysis of English? and (ii) how can the interaction of syntax, semantics, and pragmatics in different grammatical systems best be captured and explained?

http://wings.buffalo.edu/linguistics/rrg/rrg_paper.shtml

European Training Director, FOCUS, 80 Portswood Road, Southampton

Tel +2380 676928 Mob +7950 183761Fax +2380 676981 www.eurotp.org

BOOKS AVAILABLE FOR REVIEW

The following are among the books available for review by our readers. If you wish to do a book review for publication in Notes on Linguistics, contact the editor and the book will be shipped to you along with instructions for submitting the review. When you submit a review, the book is yours to keep. Contact:

Mike Cahill, Editor, Notes on Linguistics
7500 W. Camp Wisdom Road, Dallas, TX 75236
Internet e-mail: Eugenecc_Loos@sil.org. (Editorial Assistant)

Ackerman, Farrell and John Moore. 2001. Proto-properties and grammatical encoding: A correspondence theory of argument selection. Chicago: The University of Chicago Press (distributed for CSLI Publications). 197 pp. Cloth $55.95, paper $18.95.

Bauer, Laurie. 2001. Morphological productivity. Cambridge Studies in Linguistics. New York: Cambridge University Press. 245 pp. Hardback $59.95.

Bertolo, Stefano, ed. 2001 Language acquisition and learnability. New York: Cambridge University Press. 247 pp. Hardback $62.95, paper $22.95.

Börjars, Kersti and Kate Burridge. 2001. Introducing English grammar. New York: Oxford University Press. 311 pp. Cloth $65.00, paper $24.95.

Bouillon, Pierrette and Federica Busa, eds. 2001. The Language of Word Meaning. New York: Cambridge University Press. 387 pp. Hardback $69.95.

Bybee, Joan. 2001. Phonology and language use. Cambridge Studies in Linguistics. New York: Cambridge University Press. 237 pp. Hardback $59.95.

Cowie, A. P., ed. 2001. Phraseology: Theory, analysis, and applications. New York: Oxford University Press. 258 pp. Cloth $24.95.

Dekkers, Joost, et al., eds. 2000. Optimality theory: Phonology, syntax, and acquisition. New York: Oxford University Press. 635 pp. Paper $45.00.

Haspelmath, Martin. 2001. Indefinite pronouns. New York: Oxford University Press. 364 pp. Paper $29.95.

Kenstowicz, Michael, ed. 2001. Ken Hale, A life in language. Cambridge, MA: The MIT Press. 482 pp. Paper $30.00

Kristoffersen, Gjert. 2000. The phonology of Norwegian. New York: Oxford University Press. 366 pp. Cloth $99.00.

Maffi, Luisa, ed. 2001. On biocultural diversity: Linking language, knowledge, and the environment. 578 pp. Cloth $65.00, paper $34.95.

Marantz, Alec, et al, eds. 2001. Image, language, brain. Papers from the First Mind Articulation Project Symposium. Cambridge, MA: The MIT Press. 272 pp. Cloth $55.00.

Melamed, I Dan. 2001. Empirical methods for exploiting parallel Texts. Cambridge, MA: The MIT Press. 195 pp. Hardcover $32.00

Moro, Andrea. 2001. Dynamic antisymmetry. Cambridge: The MIT Press. 142 pp. Paper $20.

Mukherji, Nirmalangshu, et al., eds. 2001. The architecture of language: Noam Chomsky. New York: Oxford University Press. 89 pp. $35.00.

BEST COPY AVAILABLE

Puckett, Anita. 2000. Seldom ask, never tell: Labor and discourse in Appalachia. New York: Oxford University Press. 309 pp. Cloth $85.00.

Siegel, Jeff. 2000. Processes of language contact: Studies from Australia and the South Pacific. Montréal: Les Presses de l'Université de Montréal Distribution Fides. 326 pp.

Stockwell, Robert and Donka Minkova. 2001. English Words: History and Structure. New York: Cambridge University Press. pp. Hardback $54.95, paper $19.95.

■

THE LINGUISTICS DISCUSSION LIST

SIL has now created an email discussion list designed especially for ordinary working linguists who want to ask linguistic questions. The following are (among the) things you can do on this list:

1. Ask questions about puzzling questions you have about grammar, phonology, semantics, phonetics, or language in general.
2. Share insights about language that do not merit article-length (or even squib-length) treatment.
3. Discuss useful linguistic articles or books that you have run across (but book reviews really belong in *Notes on Linguistics* or such like).
4. Provide feedback from the fields to the SIL schools. ('I wish you had taught us about X.')
5. Provide feedback from the fields to linguistic software developers. ('I need a tool that will do X.')
6. Sell your old linguistics books that are not needed any more ☺.

Here's how the list works: You send a message (no text or subject line required) to linguisticsdiscussionlist-on@lists.sil.org. In return, you will get a 'request for confirmation' message in your email. (This is to block spammers, or to keep someone else from signing you up.) Simply reply to this message, and you'll receive a 'welcome' message, describing how to use the list, how to post to it, and (don't forget!) how to unsubscribe to it. Whenever anyone posts a message to the list, the message is 'redirected' (sort of like forwarding) to all subscribers. You can reply to any message you want, and everyone who is subscribed to the list will see your response.

All messages posted to this list are saved in an archive, which can be accessed from a web browser. The URL for the archive is https://lists.sil.org/Lists/default.html (notice the 's' at the end of 'https'). In order to view this archive, you need to log in using your email address as your name; the password is the ten-digit code which will come in your 'request for confirmation' message that you responded to. (Most web browsers can save this password, if you ask them nicely ☺, but it's probably a good idea to write down the code.)

There are presently about fifty people who subscribe to this list, many of them linguistic consultants. So if you have a difficult question, you're likely to get a good answer! **Enjoy!**

Mike Maxwell, 7809 Radin Rd., Waxhaw, NC 28173.
E-mail: mike_maxwell@sil.org

CONTENTS
(Continued)

TURKO-TATAR ETHNOGENETIC HISTORY
by Mirfatyh Z. Zakiev Yuri Tambovtsev 131

A FUNCTIONAL APPROACH TO LINGUSITIC CHANGE THROUGH LANGUAGE CONTACT: THE CASE OF SPANISH AND OTOMI by Dik Bakker and Ewald Hekking Dick Watson 136

ANNOUNCEMENTS

ADVANCED LINGUISTICS: ROLE AND REFERENCE GRAMMAR 137

BOOKS AVAILABLE FOR REVIEW 138

THE LINGUISTICS DISCUSSION LIST Michael Maxwell 139

NOTES ON *LINGUISTICS*

VOLUME 4, NUMBER 3 2001

CONTENTS

COORDINATOR'S CORNER

DICTIONARIES IN SIL FIELDWORK Michael Cahill..... 142

ARTICLES

THE IMPACT OF BILINGUAL DICTIONARIES IN MEXICAN INDIAN LANGUAGES Doris Bartholomew..... 145

LEXICOGRAPHY AND MASS PRODUCTION Ronald Moe..... 150

LEXICOGRAPHY IN THE FIELD: METHODS AND RESULTS OF THE MUNA DICTIONARY PROJECT René van den Berg..... 157

REVIEW ARTICLE

PRESUMPTIVE MEANINGS by Stephen C. Levinson William C. Mann..... 172

ABSTRACTS

PROTO–BUNGKU–TOLAKI: RECONSTRUCTION OF ITS PHONOLOGY AND ASPECTS OF ITS MORPHOSYNTAX David Mead..... 179

CONFLUENCE, CONVERGENCE, AND CONTEXTUALIZATION IN PHILOSOPHY AND LINGUISTICS: THE METHODOLOGIES OF RICOEUR AND PIKE Randolph Radney..... 180

MINHE MANGGHUER: A MIXED LANGUAGE OF THE INNER ASIAN FRONTIER Keith W. Slater..... 181

REVIEWS

LANGUAGE AND THE LEXICON: AN INTRODUCTION by David Singleton Robert D. Eaton..... 183

MODERN LEXICOGRAPHY: AN INTRODUCTION by Henri Béjoint Gilles Gravelle..... 185

MENTOR COURSES: A RESOURCE BOOK FOR TRAINER-TRAINERS by Angi Malderez and Caroline Bodoczky Lou Hohulin..... 188

(Continued on back cover)

SIL INTERNATIONAL
7500 W. CAMP WISDOM ROAD
DALLAS, TEXAS 75236 USA

CONTENTS
(Continued)

WILHELM VON HUMBOLDT: ON LANGUAGE by Michael Losonsky, ed. Alan S. Kaye 191

PRESUPPOSITIONS AND ANAPHORA by Emiel Krahmer Christoph Unger 194

ANNOUNCEMENTS

BOOKS AVAILABLE FOR REVIEW 198

EUROPEAN TRAINING PROGRAM 172

From the Linguistic Coordinator

Dictionaries in SIL fieldwork

I own a book called *Mrs. Byrne's Dictionary of Unusual, Obscure, and Preposterous Words.* In it you will find such fascinating entries as *famble* 'to stutter', *fubsy* 'short and stout', and *fussock* 'a big fat woman'. Dictionaries are a lot of work to create, but they can be fun, too!

The theme of this issue of *Notes on Linguistics* is dictionaries for minority peoples. As Betty Snell wrote me,

> There are many good reasons for making dictionaries ... when a minority people, who have often been looked down upon by the surrounding society, receive a 'real dictionary' of their language, that outside society sits up and takes notice. So do the people themselves. They discover that their language really does have verbs and nouns and a whole grammar (very complex, I might add, if it's Machi). They find out that it can be put into a book, like other peoples' languages, and passed down to children and grandchildren so they'll know something about their heritage and who they are as a people.

This was experienced firsthand by Jaap and Morina Feenstra, working with the Dogrib people of northwest Canada.[1] Until the 1980s, schooling was mostly in English. Dogrib was valued in the home but literature and literacy was virtually non–existent. Later its use was encouraged, but without much literature, motivation was low. The Feenstras talked to teachers and realized that the schools were 'screaming for a dictionary'. One of them asked Jaap, 'Don't you have a dictionary, wordlist, or ANY type of printout?'

Jaap had a simple wordlist (NOT publishable!), but planned to do a dictionary some day. At this point, he met an old acquaintance, the Superintendent of Dogrib schools, who had heard about the 'dictionary' and offered to pay the Dogrib salaries to prepare it for publication—and for the printing as well.

Jaap recognized the importance of the project: 'They wanted it NOW.' Working quickly, he and his Dogrib helpers completed 5,000 entries, some with sample sentences and idioms. They published 500 copies as a 'Preliminary Dictionary', with the understanding that the Board of Education would eventually write a more polished and complete dictionary.

---

[1] Most of the information here was originally written up in WBT Canada's 'Word Alive' periodical, June–August 1995.

BEST COPY AVAILABLE

The results were dramatic. Immediately teachers used the dictionary in schools and it was clear that the publication run had been far too small. Dogrib writing appeared on signs and billboards. Children tacked Dogrib computer printouts to the walls. Copies of the book soon began to wear out due to heavy use. Jaap comments:

> Although they spoke the vernacular, they had no confidence to write it down—until they saw it presented formally in print. ... Based on dictionary samples, they could now make an educated guess at how to spell and conjugate words not yet recorded.

Besides promoting literacy, teachers asked for Scripture portions for classroom use. The elders, aware that mass media were eroding Dogrib values, wanted to teach vernacular Bible stories, and insisted on having them in print. A friend transformed three Bible story booklets into interactive talking booklets for the computer, which were an instant hit in the schools.

SIL has a number of resources on dictionary production, originally published as the references below, available on its LinguaLinks Library CD (available through SIL's Academic Bookstore online at *www.ethnologue.com*).

In this issue of NOLx, Rene VandenBerg tells the details of how the Muna dictionary came into being. SIL's Mexico branch has a lot of experience with dictionaries, and Doris Bartholomew tells us some of the stories. Finally, Ron Moe shares an innovative approach to rapid gathering of lexical data that shows great promise for the future. (For a related approach, see Grimes 1994.) I hope you enjoy reading these as much as I did.

References

Bartholomew, Doris A. and Louise C. Schoenhals. 1983. Bilingual dictionaries for indigenous languages. Mexico: Summer Institute of Linguistics. xvii, 370 p.

Coward, David F. and Charles E. Grimes. 1994. Making dictionaries: A guide to lexicography and the Multi–Dictionary Formatter. Waxhaw, North Carolina: Summer Institute of Linguistics.

Grimes, Charles E. 1994. Mapping semantic relationships in the lexicon using lexical functions. Notes on Linguistics 66: 5–25.

Newell, Leonard E. 1995. Handbook on lexicography for Philippine and other languages. Manila: Linguistic Society of the Philippines.

Michael Cahill
International Linguistics Coordinator

■

The impact of bilingual dictionaries in Mexican Indian languages

Doris Bartholomew
SIL—Mexico Branch

Following the lead of Dow Robinson (1969), the SIL dictionaries of Mexican Indian languages were designed with the needs of newly literate communities in mind. Care is taken to find optimal entry forms for verbs and other inflected words; definitions are stated as translation equivalents in Spanish; illustrative sentences show the use of the word in a natural context both culturally and linguistically. The number of published dictionaries following this model is 29, with several more in preparation. In addition several of the earlier vocabularies have been revised and expanded. The dictionaries and grammars have had a positive impact on the language community, the academic community, and they have been important in public relations in the Mexico branch of SIL.

Impact on the language community. The existence of a bilingual dictionary and/or practical grammar either as part of the dictionary or as a separate publication is of great prestige value for the speakers of the language. For generations they have been told that they do not speak a real language because it has no grammar, dictionary or literature. A Totonac speaker heard somebody make such a remark and showed the bilingual dictionary to effectively demonstrate that Totonac is indeed a full language.

The dictionary is a treasure chest for cultural values and history. The Huave dictionary includes the old vigessimal number system, including the numeral classifiers for the numbers one to four. In current speech, numbers above four have been replaced by Spanish. When some Huave school teachers, trained by the government as ethnolinguists, saw the old number system in the dictionary they were excited and confirmed with their parents that it really was the way they used to count. Then some of them who were working on the New Testament revision insisted that chapter numbers should use the authentic Huave numerals.

The young people in many language communities are learning and using Spanish at the expense of their mother tongue. The vocabulary and the illustrative sentences in the dictionary often speak of customs and practices now unknown to the younger generation. Enrique was helping to edit the Eastern Otomi dictionary and encountered words that he did not know. Like the Huaves, he asked his parents about them. They recognized the words

immediately and told him what they meant, exactly as the dictionary entries said. He was excited that the dictionary accurately preserved these old words as well as presenting those he knows and uses.

Some of the other entries stimulated Enrique to work with two other men in a project to collect specimens of medicinal plants and write the sentences that illustrate their use in curing common ailments. The team did such a good job that the University of the State of Hidalgo published the results of their research.

The language speakers who collaborate with SIL members on a dictionary contribute from their direct knowledge of the language and provide superb illustrative sentences. Their work on the project is also a part of their own education. Several have learned to use the Shoebox computer program for managing the dictionary database. They can fill in missing information as Barcimeo did for the Guerrero Amuzgo dictionary. Work on the project had been suspended for a time and he was able to supply new illustrative sentences as needed. He also learned to do some of the editing for correctness and consistency.

A fringe benefit of a bilingual dictionary is the impact on Spanish speakers. One man who married an Atepec Zapotec woman attended the presentation of the dictionary in Atepec. He gave credit to the dictionary for helping him to better understand his wife because of the rich cultural content in the sentences.

The dictionary, especially its illustrative sentences, has proved to be an effective literacy tool. Two Otomi women were called upon to help in a Spanish editing workshop in Ixmiquilpan. They did not know how to read Otomi, though they could read Spanish. After two or three days of working through the sentences and answering questions from the Spanish editor, the women started reading the Otomi sentences for themselves. The sentences were so culturally natural and 'predictable' in content that the women easily moved into reading them. The sentences are short enough and complete enough in themselves to make easy reading material. It also helped to have the Spanish translation right there.

The bilingual dicitionaries have become important resources for native speaker professionals who are producing their own publications for use in bilingual education and in promoting the use of the written language.

A manuscript copy of the Mezquital Otomi dictionary stimulated Felipino Bernal (1986, 1996) to collect verb paradigms and later publish his book on Otomi verbs with his own copyright. The dictionary he published is a very

respectable vocabulary with grammar notes and a good number of inserts for vocabulary of a given domain. An early trilingual dictionary (Otomi–Spanish–English) is being re–edited by the Academy of Otomi Culture because so many Otomies have gone to the United States for work and need some help with English.

A printout of the Oaxaca Amuzgo dictionary provided the point of departure for a dictionary project by an Amuzgo speaker for his MA degree. He implemented some of the ideas of the SIL project, though not all, and developed some ideas of his own, producing a more professional dictionary than he might have done otherwise.

The production of dictionaries by native speakers is being promoted by several different government agencies. Some are looking to SIL for orientation and consultation for those compiling these dictionaries. Our own expertise in working with our members has proved valuable in helping these bilingual teachers.

The current interest in bilingual dictionaries in Mexico is reflected in sales to native speakers. The Mazatec dictionary and separate grammar have sold so well in the last few years that only a small inventory is left in the village. A Mixtec dictionary is now in great demand. One of the local stationery stores has been selling them to the school children and is constantly asking for a new supply.

Impact on the academic community. Our linguistic work on the field has benefited greatly from linguistic science. Dictionaries and grammars in little known languages are one way we can 'give back' to the academic community.

From the early vocabularies to the fuller dictionaries, the interest of linguists and anthropologists has been strong. A standing order list sends our publications to a good number of university libraries and individuals all over the world. Still others request individual books in their area of specialization. Our publications show up in footnotes and bibliographies in the publications of many and diverse scholars. They appreciate the reliability of our dictionaries for phonemic transcription and the accuracy of word definitions and range of meanings.

Our dictionaries and grammars have served as patterns for other scholars in their own research in Mexican Indian languages. Dr. M. took the Isthmus Zapotec vocabulary and 'filled in the blanks' for his study of a different Zapotec language, effectively 'standing on the shoulders' of the SIL team.

Dr. H. worked with a bilingual teacher in the Queretaro Otomi area to produce a dictionary parallel to the 1956 Mezquital Otomi vocabulary. He also adapted the Eastern Otomi grammar (Voigtlander and Echegoyen 1979) to fit that of Queretaro Otomi.

Our own translation teams have profited from their work on dictionary and grammar projects. A Totonac team found themselves constantly consulting their own dictionary for accurate spelling of words, especially for the elusive long vowels. They looked up a word on the Spanish side when trying to find the right word to express a concept in translating the Totonac New Testament.

Another benefit to the translation team is that the dictionary–grammar project provides a way for them to 'package' some of the things they have learned about the language and culture in the process of doing a translation program. As mentioned previously, they are often able to tap further the vast knowledge of their translation associates to accurately document the language.

Reference books prepared to help our members produce quality dictionaries have turned out to be helpful to other scholars. Louise Schoenhals' book, *A Spanish–English glossary of Mexican flora and fauna* (1988), has sold well and continues to be requested by anthropologists and linguists as well as biologists. The manual, *Bilingual dictionaries for indigenous languages* (Bartholomew and Schoenhals 1983), has helped members in other branches of SIL and recently was quoted several times by a Mexican Lexicographer (Lara 1997) as he considered and developed concepts for a theory of the monolingual dictionary. A Spanish lexicographer 'discovered' the *BDIL* while in Mexico working on the *Diccionario del Español de México*; he asked permission to translate it into Spanish because he considers it to be part of the pertinent literature for Spanish lexicography (see Carriscondo 2001).

Impact on Public Relations. The series of dictionaries and grammars has been very good for public relations in Mexico and around the world. Many know us primarily through our publications; others find out about us through the publications of those who use our materials. The web page of the Mexico branch of SIL generates many inquiries for further information. Many times we are able to oblige them.

Recently we co–published an Aztec dictionary with Madero University in Puebla, where we offer a summer course in Linguistics and Cultural Sensitivity. The public presentation at Madero University of this dictionary in a prestigious Indian language opened many doors. A prominent scholar of

Classical Aztec (and long time friend of ILV) wrote an excellent prologue to the book. There have been five other public presentations—three in universities and two in the language area. One of the speakers for the presentation at the presidential palace in Huauchinango, Puebla, is a speaker of Aztec in the neighboring state of Tlaxcala. He had worked with an American linguist on a root dictionary of Classical Aztec that was done in English. The Aztec speaker was especially impressed that the North Puebla Aztec dictionary was bilingual in Spanish and that it was presented in the language area and 'given back' to the speakers of that language.

Conclusions. Published dictionaries and grammars have an important and favorable impact on native speakers, scholars, and government officials and more. It is well worth the effort of our members to work with their language associates to prepare them and the financial investment to print them.

REFERENCES

Bartholomew, Doris A. and Louise C. Schoenhals. 1983. Bilingual dicitionaries for indigenous languages. Mexico: Instituto Lingüístico de Verano.

Bernal Pérez, Felipino. 1986. Hñähñu—recopilación de 505 verbos otomíes conjugados. Cardonal, Hidalgo: Centro Social de Cardonal, Hgo.

____. 1996. Diccionario hñähñu–español y español–hñähñu del Valle del Mezquital, Hidalgo. Ixmiquilpan, Hidalgo, México. Segunda edición 1998, Asesoría: Verónica Kugel.

Carriscondo Esquivel, Francisco Manuel. 2001. La lexicografía en las variedades no–estandar. Colección Alonso de Bonilla. Jaén (España): Universidad de Jaén.

Lara Ramos, Luis Fernando. 1997. Teoría del diccionario monolingüe. México: El Colegio de México, Centro de Estudios Lingüísticos y Literarios.

Robinson, Dow F. 1969. Manual for bilingual dictionaries. 3 vols. Santa Ana, CA: Summer Institute of Linguistics.

Schoenhals, Louise C. 1988. A Spanish–English glossary of Mexican flora and fauna. Mexico: Instituto Lingüístico de Verano.

Voigltander, Katherine y Artemisa Echegoyen. 1979. Luces contemporaneas del otomí: Gramática del otomí de la Sierra. Gramáticas de Lenguas Indígenas de Mexico 1. Mexico: Instituto Lingüístico de Verano.

Doris Bartholomew, P. O. Box 8987, Catalina, AZ 85738-0987.
E-mail: doris_bartholomew@sil.org.

Lexicography and mass production

Ronald Moe
SIL—Linguistics Consultant, Uganda Tanzania Branch

Lexicography is a fruitful area for mass production techniques. Consultants can provide templates (standardized guidelines) to enable people who do not have a specialty in lexicography to produce extensive, quality dictionaries. Using an exhaustive list of semantic domains, it is possible to collect over 10,000 words[1] within two weeks. Doing so at the beginning of a project provides a substantial dictionary which can be used throughout the project for a variety of tasks. Collecting words by semantic domain results in a dictionary that is classified by domain. It is far more efficient to expand the dictionary field by field than word by word. It is far more insightful to investigate semantics domain by domain than word by word.

Mass production. Mass production techniques are highly effective whenever large numbers are involved. With thousands of languages in the world and tens of thousands of words in each, we need to collect and describe something on the order of 100,000,000 words. Efficiency is not a luxury but a necessity. Large publishing houses can afford to hire scores of professional lexicographers to work on a single language. SIL cannot. However, mass production techniques, such as the use of templates and task specialization, can multiply our efficiency and productivity enabling mother tongue speakers to produce massive dictionaries quickly and with relatively little training or consultant help.

Within SIL most dictionaries are produced one word at a time over the course of a project. Words are collected as they are encountered. A word may be researched and described when it is added to the dictionary, or it may be entered with very little description. The result of such a hit or miss approach is usually a small dictionary that is very uneven in its breadth and depth of coverage. We have policy documents that recommend that a team should collect 1,500 words before beginning translation. Considering that the Greek New Testament contains 12,000 words, a database of 1,500 is going to be insufficient to suggest translation options.

---

[1]In the interests of simplicity and naturalness, if not accuracy, this article employs the term 'word' to refer to lexical items of all sorts, including roots, derivatives, compounds, idioms, and phrases.

BEST COPY AVAILABLE

Many teams do not do serious work on the dictionary until after the translation is completed. This is tragic, because the dictionary can be a tremendous aid to translation and many other aspects of a program, including language learning. Virtually every computer program SIL uses for language work depends on a dictionary database, including those for interlinearization and phonemic analysis. I recommend that at least 12,000 words be collected in a two–week workshop at the very beginning of the project, so that the dictionary can serve as a tool throughout the project.

Templates. The Bantu Initiative has recommended the production of templates to facilitate various tasks. A template is a mold or pattern that can be used repeatedly to produce similar objects. A ruler is a simple template for drawing straight lines. An example of a linguistic template would be a guide to producing a phonology statement. Much of the work in SIL could be facilitated by the use of templates. Whether producing a dictionary, orthography guide, or grammar, consultants can write guides and outlines of various sorts to standardize and facilitate the procedure and product. Consultants can design linguistically universal templates or modify a universal template for use within a particular language family.

SIL's training tends to focus on abstract principles to guide the researcher. The standard works on lexicography are no exception, giving principles along with examples to illustrate how the principle is to be applied. But it takes a great deal of study, imitation, practice, and correction to learn how to apply abstract principles to real life. A template can shortcut much of this process, enabling a person with much less training to produce a quality product, because he is guided and restricted. A lack of constraints in lexicography results in very messy databases. Constraints may hinder a professional craftsman, but even craftsmen use templates, and they are incredibly helpful to someone who does not know all the intricacies of the job. With a ruler, even a child can draw a straight line.

A template also makes the process accessible to more people. As long as the procedure remains a complicated technique in the mind of one person, it is limited to the time he has available to devote to the task. A dictionary template enables people who do not have a specialty in lexicography to produce extensive, quality dictionaries. It enables the speakers of a language to do the bulk of the work and places the development of their language in their hands.

Lexical relations and semantic domains. The words of a language are organized in the mind in a multi–dimensional network of relationships of various sorts. Words are linked by patterns of syntactic distribution (part of speech), phonology (e.g. rhyme), semantics, and pragmatics.

BEST COPY AVAILABLE

Lexicographers have catalogued many types of semantic links, called lexical relations (or lexical functions). Lexical relations tend to cluster around a central nexus. The nexus may be an area of life or some general notion. One definition of a semantic domain is 'an important idea and the words directly related to it'. For instance the Generic–Specific lexical relation is the basis for many semantic domains, e.g. 'Walk' (Generic) and *stroll, amble, strut* (Specifics). Other domains such as 'Sleep' center around an area of life and include words such as *snore, pajamas, bedtime,* which are related to *sleep* by various lexical relations.

Many lexicographers have recommended that we utilize semantic domains and lexical relations to elicit and investigate words. What has been lacking is an exhaustive and universal list of domains. For instance, the *Outline of Cultural Materials* (Murdock, et al, 1987) presents a list of anthropological domains but is missing many lexical domains. *Roget's Thesaurus* (Roget, 1958) has 1000 domains, but due to its purpose it also omits many domains. Louw and Nida (1989: xix) admit that their list is uneven due to the subject matter of the New Testament. To fill this gap, I have been developing a list of semantic domains and related materials to facilitate the production of dictionaries. The list can be used to elicit, classify, and investigate the words of a language. Once a dictionary is classified by semantic domain, it can be sorted by domain for research purposes, to produce a semantic index, or to publish a semantically organized dictionary such as Louw and Nida's Greek–English lexicon.

Every effort has been made to make the list as exhaustive as possible so that every lexeme in every language can be reasonably classified under one of the domains. Semantically complex lexemes are a combination of simpler or more general notions and can be classified under several domains. For instance *shrug* 'to raise the shoulders to express uncertainty or indifference' could be classified with *shoulder* under 'Body—Parts of the body', under 'Movement—Moving a part of the body', or under 'Communication—Gestures'. The list will undoubtedly require refinement as we gain experience using it with languages.

As I have compared lists of semantic domains from around the world, it has become clear that almost all of the domains are universal. The differences come from minor differences of culture and the necessity to squash the multi–dimensional network into a two–dimensional list. Even organizing the list hierarchically fails to maintain all the semantic links. Some links can be maintained, while others must be lost. I have also tried to attain a level of detail such that each domain would contain ten to twenty words. At this level of detail the list contains approximately 1500 domains so the

organization of the list is etic, somewhat arbitrary, and based on the commonalities of the lists available to me.

Collecting words. Eliciting vocabulary has been a topic of interest for some time, and the literature contains a wealth of practical suggestions, such as using semantic domains and concording a text corpus (Beekman, 1968, Ballard, 1968, Pallesen, 1970). However, the use of semantic domains is by far the most effective, efficient, and productive. The combination of semantic domains and lexical relations is particularly powerful. Since the entire lexicon is tied together by lexical relations, the mind can jump rapidly from word to word, especially within a domain.

An extensive list of semantic domains fleshed out by lexical relations makes it possible to elicit efficiently a large percentage of the vocabulary of a language in a short time. This sort of systematic approach will ensure that the dictionary covers nearly all domains of the language and does so to a relatively uniform depth.

Although lexical relations are very helpful in thinking of lexemes, the number of lexical relations is quite large. It is very inefficient for someone to have to think through the entire list for each new word encountered. Rather than requiring each lexicographer to reinvent the wheel, I have thought through each domain, identifying the lexical relations applicable to that domain. It is more efficient for a single consultant to think through the theoretical issues than for each end user.

Lexical relations are easy to use but hard to grasp in the abstract so I have worded each relation in the form of a simple question. For example, the domain 'Wind' has the following productive lexical relations:

> What words describe a wind that lasts for a short time? *breath of air, puff of wind, gust*
>
> What words describe a light wind? *draft, breeze*
>
> What words describe a strong wind? *gale, howling (wind)*
>
> What does the wind do? *blow, freshen, rise, fan (flames)*
>
> What words describe the direction of the wind? *north wind, northeaster, updraft*
>
> What sounds does the wind make? *sigh, moan, whistle, howl, shriek*

Answering these questions elicits a wealth of lexical material which might otherwise be overlooked. The example words following each question are

merely meant to be illustrative. It takes very little mental effort to think of other words.

Example of a domain template:

> **Lose consciousness.** Use this domain for words related to losing consciousness, including fainting, being knocked out, and anesthésia.
>
> For visions, hallucinations, and spiritually induced trances use 'Vision, hallucination'.
>
> What words refer to losing consciousness? *lose consciousness, go unconscious, faint, swoon, pass out, black out, be knocked out, go into a coma*
>
> What words refer to the state of being unconscious? *be unconscious, be in a coma, be out, fainting spell*
>
> What words refer to something causing someone to lose consciousness? *knock (someone) out, put under (anesthesia)*
>
> What words refer to regaining consciousness? *regain consciousness, come to, come out of (the coma)*
>
> What causes someone to lose consciousness? *hit on head, be sick, pain, shock, anesthesia*
>
> What happens or what symptoms occur when someone begins to lose consciousness? *feel faint, feel dizzy, stagger, become incoherent*

Elicitation workshop. The elicitation procedure was tested using a beta version of the semantic domains list in a workshop for the Lugwere language of Uganda.[2] In ten days, fifteen participants collected over 10,000 words and 1000 example sentences.[3] One participant said, 'The words are falling out of my head.' With a little practice, a person can collect words almost as fast as he can write. I am currently revising the materials and method so that even better results should be possible. The procedure consists of the following:

1. Organize a workshop with twelve to twenty participants. The participants need to be educated, literate, and bilingual so that they can read the domain templates and write down the vernacular words that belong to each domain.

---

[2] Thanks are due the Bantu Initiative for funding this workshop and initial research for the semantic domain list.

[3] By comparison many dictionaries are published with only 3000-5000 entries.

2. Print the list of domains—one domain per page. Divide the domains into sections of about twenty domains and place each section in a folder.

3. Give instructions to the participants explaining the concept of a domain, the procedure, and working through some domains as a group to give them some practice. Decide on citation forms for the major parts of speech. Try to minimize orthography issues.

4. Divide into small groups. Allow each group to choose a folder of interest to them. Try to get people to work on domains that they have special knowledge of (pastors work on 'Religion', farmers work on 'Agriculture'). Have each group review the domains in the folder then work on one domain at a time. Have them read the instructions then think of as many words as they can. Note that many of the English examples in the templates given above are multi–word lexical items. The participants should be encouraged to think of idioms and phrases, and cautioned against literally translating the English phrases. The phrases they think of will either be cited in the dictionary as full entries or included in an entry as an example of usage.

5. If time permits, have them go back and give a simple gloss in the national language. If there are unpredictable word classes, have them give the plural of nouns and affixed forms of verbs that will enable identification of the class. However, the primary goal is to collect as many words as possible. These other tasks can be done later.

The workshop facilitator should closely monitor each group's work—especially at the beginning. Incorrect work will result in a huge cleanup job later.

In the Lugwere workshop the participants were highly motivated and did not get bored, partly because the topic was constantly changing. They finished going through all the domains in about five and a half days. The participants improved over the course of the workshop, so that by the end they were coming up with twice as many words per domain. The last few days I asked them to review the earlier domains that had few words and see if they could add more words.

The list of domains is a tool that can be used creatively in a variety of situations to achieve a variety of goals. If you are working in a monolingual situation, you can use the domain templates as a tool for one–on–one elicitation. One language learning goal could be to learn one or two high–

BEST COPY AVAILABLE

frequency words from each domain or to concentrate on those domains which you frequently deal with.

Expand the dictionary field by field. Once the words of a language have been collected, the dictionary can be filled out field by field. It is far more efficient to expand the dictionary field by field than word by word. Some fields can be added using macros. Speakers of the language can be trained to do things like add the part of speech or write example sentences. Some tasks, such as correctly identifying all the parts of speech, require a trained linguist, but the vast majority of the work can be done by non–linguists.

Investigate semantics domain by domain. One benefit of using semantic domains to collect words is that the resulting dictionary is automatically classified by domain. Lexicographers recommend that words be investigated in semantic sets. It is far more insightful to investigate words domain by domain than in isolation. Definitions can be standardized within a domain. It is also better to write example sentences for all the words of a domain at one time. Many pragmatic issues, such as connotation and register, are also better investigated within the confines of a domain. Anthropological issues also tend to be domain specific.

Version 2 of the domains list will include instructions for investigating words. Each domain template will note the semantic features which are likely to distinguish the members of the domain and give sample definitions. It will note anthropological, pragmatic, and other potential issues. Version 1 of the list of semantic domains and instructions for its use in eliciting words will be available March 1, 2002.

REFERENCES

Beekman, John. 1968. Eliciting vocabulary, meaning, and collocations. NT 29.

Ballard, Jr., D. Lee. 1968. Studying the receptor language lexicon. NT 29.

Louw, Johannes P. and Eugene A. Nida. 1989. Greek–English lexicon of the New Testament: based on semantic domains, Vol. 1. New York: United Bible Societies.

Murdock, George P., Clellan S. Ford, Alfred E. Hudson, Raymond Kennedy, Leo W. Simmons, and John W. M. Whiting. 1987. Outline of cultural materials. 5th ed. New Haven: Human Relations Area Files.

Pallesen, Kemp. 1970. More on elicitation. NT 36.

Roget, Peter Mark. 1958. Roget's Thesaurus. Harmondsworth, Middlesex: Penguin Books.

Ronald Moe, P.O. Box 44456, 00100 Nairobi, Kenya. E–mail: ron_moe@sil.org

Lexicography in the field: Methods and results of the Muna dictionary project*

René van den Berg
Indonesia Branch and Asia—SIL, Darwin

1. Introduction. In this paper I report on a specific activity in which I have been engaged over the last number of years, namely the making of a dictionary. This project has resulted in the publication of a 709-page Muna–English dictionary (van den Berg 1996) and a 451–page Muna–Indonesian dictionary (van den Berg and La Ode Sidu 2000). This contribution, therefore, is very practical: to report on how this particular project was initiated, how the data were obtained, how the data were processed, and what decisions had to be made before and during the actual project. I believe the Muna dictionary project has been quite successful and offer here as much background information as possible in the hope that a similar approach might be useful for other field workers. Each language project is unique, but lessons learned during this project can be used profitably in the process of documenting other minority languages.

2. Muna: background information. Muna is one of the large islands situated off the southeast coast of Sulawesi, Indonesia. It is approximately 110 km long and 50 to 60 km broad. Muna is a fairly dry coral island with no real mountains and hardly any rivers. The staple crop is maize, with occasional rice, supplemented by yams, tubers, and sweet potatoes.

The population of Muna is over 200,000. All the inhabitants speak one language, also called Muna, an Austronesian language. It is also spoken on the west coast of the neighboring island Buton. The total number of people speaking the language is probably around 250,000–300,000. The major dialect division is between the prestigious northern dialect and the more peripheral southern dialect, which is only 87 percent cognate with the north. Most people under 40 years of age in Indonesian are bilingual to various degrees of proficiency. A grammatical description of Muna is found in Van den Berg (1989).

---

* This article is an abridged and updated version of a paper given at the KITLV workshop on '(The study of) Endangered languages and literatures in Southeast Asia', Leiden, December 1996. A preliminary version was presented at SIL–UK in the summer of 1996.

The majority religion on Muna is Islam (98 percent), although pre–Islamic beliefs and practices are still widespread. A few villages in the southern part of the island are predominantly Catholic. These Christians speak the southern dialect.

Muna does not have a written tradition and books in the local language are all of recent date. Apart from linguistic works, there are a number of 'popular' books such as a spelling guide, a trilingual conversation book, a riddle book, some reading books, trial Scripture publications, and school books for lower and middle schools. The language is enjoying increasing prestige because of its documentation and role in the schools.

3. The stages of the dictionary project. The Muna dictionary project has gone through two stages. The first stage began during our first field work period on the island in 1985–86. During that period my wife, Lydia, and I stayed in the district capital, Raha, where we focused on grammatical analysis and text collection (in preparation for publishing the grammar). However, a handwritten dictionary file was begun and regularly updated, stored on handcut file cards in a wooden box which ultimately contained some 2,200 main entries. This phase ended a few years later in 1990 when all that material was keyboarded onto computer in Ujung Pandang (now called Makassar).

The second stage began in the autumn of 1992 when our family was allowed to return to Muna for a new field work period until April 1994. We lived in Watuputih, a village some five km from Raha. In addition to translation work, a comprehensive trilingual dictionary project was one of the components of this field period. My main concern was to acquire as much material as possible within that time frame (only 18 months) and to make optimal use of local expertise. Fortunately, my official counterpart at the Haluoleo University in Kendari was La Ode Sidu, M.S., a lecturer in Indonesian and a native speaker of Muna with a keen interest in lexicography. Even though the distance between the provincial capital Kendari and Watuputih was considerable, we were able to set up a system of cooperation and communication that worked well.

Using different methods which I will explain below, we acquired new material. This new material was immediately entered on computer. La Ode Sidu worked independently in Kendari and his target was to cover one letter a month, after which the material (on diskette) was sent to me in Watuputih. I went over his material very carefully, adding some English and noting further questions or points of discussion. A most profitable part of the process was to go over the material again with an extremely knowledgeable language helper named La Ada. For almost a year and a half, this former

teacher came to our house every night to work on the dictionary. Many more new words and meanings were detected, and both of us thoroughly enjoyed the work. During this process I used Shoebox 2.0, a computer programme developed by SIL for doing lexicography.[1]

Trial print–outs were made of entries for sevéral letters which were then checked by other speakers of Muna. Upon our return to Holland at the end of April 1994, I started working more systematically on adding English glosses and translations, as well as weeding out errors and inconsistencies.

This stage was finished in the autumn of 1995 in England, where we had moved earlier that year. At that point I concentrated on preparing the Muna–English version for official publication, while the Muna–Indonesian edition was only produced in temporary format. I found that we needed four rounds of corrections and proofreading of the Muna–English edition. Just before the end we also produced a reverse index English–Muna. The speed which with the book was typeset, printed and bound may well be a record in the history of the KITLV Press—just over three months. It came off the press in the middle of April 1996, just a week before I made a long–planned trip to Indonesia, and consequently I was able to present the official Muna–English dictionary and the preliminary Muna–Indonesian edition at two formal ceremonies: to the rector of Universitas Haluoleo in Kendari and also to the *bupati* (district head) of Muna and the head of the education office in Raha.

This whole project could not have been completed without the help of many. They are recognized on the last page of the introduction to the dictionary .

4. Preliminary decisions. Any dictionary project requires that considerable time be spent thinking about important preliminary questions. Sources found useful in this respect before and during the project were Bartholomew and Schoenhals (1983) and Newell (1986), the latter in its final version as Newell (1995). After finishing my own project I also looked at the excellent works of Al-Kasimi (1977) and Svensén (1993), none of which were available to me on the field. Preliminary questions are important because the answers to them determine both the process and the final outcome. In the case of the Muna project the following factors needed to be addressed.

- Priority and time. How important is it? How much time is available? We felt that after the grammar, the next project should naturally be a

---

[1] Editor's note: The Linguist's Shoebox version 5 is now available for purchase on CD-ROM at http://www.ethnologue.com/tools_docs/showbox.asp.

good comprehensive dictionary. Because of the contract situation between SIL and the Indonesian National Institute of Science (LIPI) only 18 months were available on the field. This meant that our time was limited and that good planning and optimal use of resources was essential.

- Audience. We felt that our target audience was two–fold: first of all the Muna people, as a documentation of their language. Users may want to consult a dictionary for matters of spelling, word classes, Indonesian equivalents, or just to browse in and discover the riches of their own language. All this would be especially valuable for teachers. Indonesian should therefore be the target language.

  The second audience we perceived to be outside of Indonesia: researchers such as linguists and anthropologists. For them an English target language would be necessary. At the same time such a dictionary would put the language on an official, recognized level by raising its status and prestige. It was therefore decided that a trilingual dictionary (Muna–Indonesian–English) would be our target. However, the end result was two bilingual dictionaries, as will be explained below.

- Funding and co-workers. Earlier attempts to obtain funding for a dictionary project had failed in 1989. As a result, **money was scarce** and I was unable to carry out my original plans to employ a team of co–workers and set up an office. Instead, I could only employ three part–time co–workers, two of whom were introduced earlier:

  La Ode Sidu in Kendari, my official counterpart. I loaned him a computer and gave him some basic training in using it for our purposes; he was paid on a monthly basis. He is responsible for much of the Muna material and for most of the Indonesian translations.

  La Ada, an extremely able elderly man who also lived in Watuputih; he became a close friend and our main source of information. Surprisingly, he refused to be paid.

  Syahruddin, a recent university graduate who lived in the nearby village of Dana. He wrote many of the semantic field essays.

  After our return to the Netherlands in 1994 I made another attempt at obtaining funds, this time for the final English checking, proofreading, and publication. Again, I was unsuccessful. Thanks to our move to England and our stay on the Wycliffe Centre there, we found enough volunteers to help with the English check and the proofreading.

Fortunately no further funding was needed for the Muna–English publication once the KITLV Press had accepted the manuscript.

All in all, the project has run on a very low and limited budget, mainly drawn from personal funds. Looking back now, it seems a miracle in more than one way that it ever materialized.

- Spelling. This issue had already been resolved by the Muna language team in 1991, the results of which are reflected in the *Pedoman Ejaan Bahasa Daerah Muna* (Muna spelling guide). As much as possible Muna spelling conforms to Indonesian, with a few exceptions for sounds that are not present in Indonesian, e.g. <gh> stands for a voiced uvular fricative, as in *ghome* 'to wash'.

- Dialects. The northern dialect is prestigious and forms the basis of the grammar, the spelling guide and of the current school material being developed. It was therefore a natural decision to limit the project to this variety. Another factor which facilitated the choice of dialect was that La Ode Sidu was born in Watuputih, the village where we lived during the project, located in the northern part of the island.

- The role of the computer. Working with a CAF laptop (368SX, 60MB) on the field, I first entered the data simply using standard field markers in Word–for–DOS. However, searches took a long time, alphabetization was not automatic and finally I was persuaded to try Shoebox (version 2.0). That was a real help, although as a simple computer user, I still needed expert help with creating print files in order to make publishable printouts. The computer helped the project in several other ways. First, an Interactive Concordance (IC) programme created lists of words or roots as found in texts. Secondly, the computer created a blank Muna dictionary (see below). Creating the reverse index could also not have been done without the computer. Finally, proofreading and typesetting were also greatly facilitated by the computer.

- Publisher. While I was planning the second phase of the project in late 1992, I contacted the KITLV Press in Leiden to find out if they were interested in publishing a Muna–Indonesian–English dictionary. Since a trilingual dictionary of the neighboring Wolio languge had already appeared in 1987 there was at least a precedent. Dr Poeze, the director of the KITLV Press, agreed in principle to publishing a Muna dictionary, but only to a Muna–English edition for practical reasons (size and cost). A Muna–Indonesian edition would have to be published separately, preferably in Indonesia. Since the current Muna–English

edition is 709 pages and costs DFL 125 (around $60), this decision seems to have been well–founded.

5. Data collection: how do you get what you want? The first concern of a good lexicographer is to get adequate material. When you start from scratch this need is all the more pressing. Two Indonesian sources were of limited value, but in combination with our own handwritten database there was at least a base to begin work from. We acquired our lexical material by means of the following three methods:

a. From oral literature texts, both prose and poetry. During the first phase (1985–86) we had collected some 75 texts. Most of these were written by native speakers, while others were recorded live and transcribed later. At the end of phase two we had some 130 texts of various genres and of different lengths, such as 'The tortoise and the monkey', 'My experiences in teaching English', and 'The inauguration oath of the King of Muna'.

These texts were processed by reading them carefully and noting down new vocabulary. New words or new meanings were entered on new cards (during phase one) or into the computer (later). After they had been keyboarded, these texts also served as the basis for making a computerized concordance, using the Interactive Concordance programme mentioned above. The following example shows part of such a concordance on three roots: *gholi* 'to buy'; *ghohi* 'to tell a lie', and *ghondo* 'to look'. The first number refers to the text number, the second to the line.

gholi

9 44: pasina noalamo \se44. Gholino kansuru negholiane mbololo
9 44: Gholino kansuru negholiane mbololo \se45. Pada aitu
36 9: \se9. Barangka noala negholianemo kahitela bhahi o mafusau
36 13: Miina nakodoi so naegholighoo kaago sigaahano \se14.

ghohi

13 29: pikore \se29. Gara neghohi \se30. Dadi o karambau notalomo
40 230: we panda watu deghohi \se231. Ane ta miehi andohi
40 265: ndoke \se265. Omeghohiimu ihintuumu itu \se266. Ingka
43 81: isahihino \se81. Kaghohindo \se82. Dofolo-fololuhi a

ghondo

1 40: ndokea Soba dokala doghondoegho Dosikalahamo \se41.
1 68: kapoluka tora Tameghondohighoomo wutomu \se69. pata
3 5: sau melangke noghoo–ghondomo fatowalae so neweino maitu
6 12: sikola ingka ama \se12. Ghondofaanda aihimu itu \se13.
10 27: aitu adhi–adhinia noghondohimo tora apa–paando \se28.

Such a concordance is helpful in several ways. It facilitates finding example sentences, recurrent collocations of the word, the most frequent and regular affixation of the root, and anything unusual which may have gone unnoticed. Since this IC programme was slow (on my computer) and could not run concurrently with Shoebox, processing took longer than I expected and consequently this method was used only for a limited number of items. Ideally one should have a concordance of the whole corpus, although it is easy to get drowned in the data.

b. Although collecting such texts is worthwhile, one quickly discovers that the number of unknown words in a new text diminishes considerably with time. I therefore employed a second method, which I call the 'semantic domains approach'. I drew up a list of some seventy topics and had my assistant, Syahruddin, write lists of words or prose essays about these. Topics included:

kinship terms
the loom
diseases
birds and fish
feelings
house parts
clothes
furniture
ornaments
making fire
food crops
ways of preparing food
food and drink containers
ways of preparing fish
making palm wine
getting and storing water
children's games
musical instruments

When he had written an essay, I would read it, underline any new or unknown words and then ask him to provide an Indonesian equivalent or explanation. This was an excellent way of acquiring new material, as well as learning more about the culture.

c. The third method was the so–called blank dictionary (or dictionary shell). The suggestion for using a blank dictionary comes from Caroll (1966), who used it as an elicitation technique in Polynesian lexicography. This technique provides the lexicographer with a computer–generated list of all possible disyllabic word shapes in the language, to be checked by native speakers as to whether or not these words actually occur. Since syllable structure in Muna is very simple (only V and CV), this method worked particularly well. Barbara Friberg, a colleague in Sulawesi, made the computer programme and a print–out, consisting of some 19,000 possible words, printed in five columns on 91 pages. Words that actually do exist as bases were then

circled by speakers of the language. La Ode Sidu had his own list in Kendari from which he worked, but our village friend La Ada was especially good at this.

The next step is to elicit the meaning of the circled words if not yet known. By careful use of this method, the lexicographer can be sure he obtains at least close to 100 perecent of all disyllabic roots in his dictionary. Since these do constitute between 50 and 60 percent of all roots in Muna, one can be reasonably sure that in combination with the other data the resulting dictionary is a good sample of the lexical richness of the language. Roots of three or more syllables have to be found using methods 1 or 2, although La Ada sometimes added syllables when longer words came to mind. The non–existent root *ghoghe*, for example, triggered the word *ghoghea* 'dry leaf of coconut palm'.

6. Organization of entries: what information should be included? Another big issue was the actual organization of the database, as illustrated by the following questions.

a. What should the basic arrangement of the data be? We decided to follow the common practice in Austronesian lexicography of presenting entries in root arrangement. The advantage is that all related words come under the same entry, but the obvious disadvantage is that a word needs to be stripped of its affixes before the dictionary user can locate it. In a morphologically complex language like Muna this requires considerable grammatical sophistication, not only for the native speaker but also for the foreign researcher, who first has to learn about some basic morphological processes before he knows that *dakumalamo* 'we will go' is to be found under *kala* 'go'. I devoted considerable space in the introduction to the dictionary to this process of affix–stripping.

b. How much grammatical information should there be? We decided to give part of speech abbreviations for all entries (nouns, verbs, numerals, etc.), with subdivisions for verbs into transitive, intransitive, and stative verbs and also for their morphological classes: *a-*, *ae-*, or *ao-*. The introduction therefore explains Muna verbal morphology in quite some detail.

c. How much productive morphology should be included? Since Muna has a very rich morphology, it would be very impractical to give all possible word shapes. On the basis of *gholi* 'to buy', for example, the following list is completely predictable, and this is only part of the first person singular!

BEST COPY AVAILABLE

- *aegholi* 'I buy/bought (an indefinite object)'
- *agholi* 'I buy/bought (a definite object)'
- *agholie* 'I buy/bought it'
- *aegholiane* 'I buy/bought (an indef. object) for her/him'
- *agholiane* 'I buy/bought (a def. object) for her/him'
- *aegholighoo* 'I buy/bought (an indef. object) for someone'
- *aegholimo* 'I already bought (an indef. object)'
- *agholiemo* 'I already bought it'
- *aegholianemo* 'I already bought (an indef. obj.) for her/him'
- *aegholighoomo* 'I already bought (an indef. obj.) for someone'
- *aghumolie* 'I will buy it'
- *aghumoliane* 'I will buy it for her/him'

It is obvious that including all these words is both impractical and undesirable. Therefore some familiarity with the morphology of the language is assumed (this is treated in the introduction) and only some derivational affixes are treated as subentries. Under *gholi*, for example, one finds *gholifi* 'buy (many items), shop'. This derivation is included because of the unpredictable thematic consonant *f* in the suffix *-fi*. Another derivation listed there is *mansogholi* 'buy anything', the affix *manso-* is not productive on transitive verbs. Everything else (literally hundreds of derived forms!) is predictable from the fact that *gholi* is a transitive verb of the *ae*–class.

d. Should example sentences be included? We decided illustrative examples are very important, because they enhance understanding, show the word in a live context, reinforce the grammatical information, and provide an opportunity for showing cultural information. For some verbs and nouns we used more than one example sentence in which the second contains another form or shows another sense. For example, under *ala* 'take, get', the first example sentence is *Dakumala daeala sau* 'Let us go and get wood'. The second is *Sau hae nealamu itu?* 'What kind of wood have you taken?' In this instance, the forms *daeala* (first person plural irrealis) and *nealamu* (passive participle with second person agent) illustrate the regular inflections of *ala*. Some of the illustrative examples have come from texts, but the majority were provided by La Ode Sidu and La Ada as they worked on the entries.

e. Should we include pictorial illustrations? At first this was not really considered since we did not have an artist on the team. However, it turned out that La Ada had great skill in drawing and often made useful sketch drawings during elicitation sessions to illustrate words. As a result I gave him a list of words for which he drew pictorial illustrations. These are included in an appendix of the dictionary. Although some of

BEST COPY AVAILABLE

them are rough and not drawn to scale, they do give an adequate idea of the shapes of many of the lesser–known artifacts, such as tools and weapons, musical instruments, the loom, the house, containers, ornaments, fish traps, etc. Many of these items are now obsolete on Muna which obviously increases the documentary value of the dictionary.

f. Should loanwords be included? The general feeling of my co–workers was that Indonesian loans should be excluded since they are not 'true' Muna. However, many loans have undergone changes in sound or meaning and are an integral part of the language. I therefore decided to include them generously and consequently one will find words such as *arilodhi* 'wristwatch' (through Indonesian *arloji* from Dutch *horloge*) and *banara* 'right, true' (from Indonesian *benar*). I only made exceptions for very recent unadapted borrowings from Indonesian. In most cases the Malay/Indonesian origin of such words has been indicated at the very end of the entry in square brackets.

g. How much cultural information should be presented? I strongly believe that a good dictionary is not just a list of words but acts as a window on the culture offering interesting explanations and insightful illustrative examples. Thus under *karia* 'puberty ritual for girls' one finds the following extra information:

> They are secluded in a dark, closed room for about four days and nights with very little food (*kaghombo*); afterwards they are dressed up and adorned, then led out of the house (*kafosampu*) to a platform where they sit in a row and are prayed over and touched with dirt; finally each one in turn performs a short dance during which presents are thrown to them (*kaghoro*). After a week the *kafolantono bhansa* ceremony is held, to find out who will marry soon.

This is exceptionally long and clearly oversteps the limits for a traditional dictionary. However, without an existing proper Muna ethnography, such information is in my view not redundant and definitely captures the meaning and spirit of the *karia* festival.

h. Is a reverse index needed? The answer is clearly yes, for the benefit of both native speakers and foreign researchers who want to find the Muna equivalent of an Indonesian or English word. Here the electronic age brings its blessings by making an automatic reversal possible, although some work remained to be done beforehand. Notice that a reverse index is nothing more than a wordlist; it is definitely not meant to be a full–fledged English–Muna dictionary.

Consider the following example. There are at least seven words for 'lice' in Muna, four of which are listed below:

- *kahama-hama* 'very small louse'
- *kaghughi* 'small chicken louse'
- *out* 'louse'
- *tuma* 'clothes louse'

Some of the reversal process had to be done manually, such as pruning modifying words and putting the most general equivalent first (*otu*). The final result is as follows:

louse *otu, kaghughi, kahama–hama, tuma*

i. What should the final result look like and in which format should all the diverse information be presented? There are of course dozens of possibilities, but in consultation with the director of the KITLV Press we decided to print main entries bold and capitalized, to print subentries (derivations, compounds) bold but not capitalized on a new line in alphabetical order, and to print example sentences in italics. No columns were used, as we anticipated problems with long words. These would all have to have been broken off manually.

7. Difficulties. It may be useful in this section to look briefly at a few areas where I experienced difficulties. The first was the most persistent one: the difficulty of finding correct equivalents in English and Indonesian. This does not only apply to technical terminology, but also to what appear to be simple concepts such as bodily positions. For example, there are many Muna words corresponding to 'to sit', and native speakers were often unable to provide me with Indonesian equivalents, but rather acted it out. Here are some illustrations (with Indonesian equivalents):

- *ngkora* 'sit' (generic) [Ind *duduk*]
- *kapala* 'sit (on something raised, with legs downwards, not on the floor)' [Ind *duduk (dengan kaki ke bawah)*]
- *ghimpo* '(of women) sit politely on the floor with the legs bent to one side' [Ind *duduk bersimpuh*]
- *miminsoro* 'sit on the floor with outstretched legs (considered very impolite)' [Ind *duduk dengan kaki lurus ke depan*]
- *tangkughase* 'sit with the hands under the chin' [Ind *duduk sambil menopang dagu*]

It takes patience, persistence, and a good command of the target language to come up with these definitions, including their cultural connotations.

With technical vocabulary this gets more difficult. In the case of flora and fauna the expert help I needed was simply not available, and with only a few books on fish and birds and nothing on trees, little more could be done than say 'k.o. fish' or 'k.o. tree' (k.o. = kind of).

Only in two selected areas have I made a conscious effort to obtain technical vocabulary: house–building and the loom. Since I had obtained detailed illustrations, I was able to work on this in England. As a result, one finds construction terminology such as *dolo* 'purlin; horizontal beam which supports the rafters' and *garaga* 'strut beam to stiffen the floor joists'.

Another very difficult but interesting area of vocabulary is onomatapoiea. Like other languages in the area, Muna has dozens of words describing various sounds. Again, attempting to find a correct translation equivalent was a major undertaking. Consider the following examples:

bheghu 'sound of mangoes falling in mud or slapping a full stomach'

dhepa 'sound of pillow falling on the ground, or flapping wings of fighting cocks'

kagha 'kind of choking sound (as when holding a cat by its neck)'

Given adequate resources and better communication channels (especially time and expert help through e–mail) some of these difficulties could have been overcome. However, I personally view this dictionary as only the first important step in Muna lexicography. It is my hope that native speakers will build on it and make refinements and additions leading to a revised edition. However, Dr Poeze informed me that he considered this the only published Muna dictionary for both the 20th and the 21st century!

A last difficulty that may be mentioned concerns conflicting information from native speakers. It is not uncommon, especially for low–frequency words, to have native speakers disagree on what a certain word means. In such cases La Ada and La Ode Sidu have always had the last say; in the few cases they disagreed I have listed both meanings or forms.

8. Reception and Indonesian version. Of the Muna-English dictionary only 200 copies were printed. To date, some 85 have been sold. Several scholarly reviews have appeared (by Bernard Comrie in *Linguistics*; by Mark Donohue in *Language*, by Robert Blust in *Oceanic Linguistics* and by Bernd Nothofer in *Bulletin of SOAS*)—all favorable.

When the Muna–English publication came out, I also made a preliminary 480–page print–out of the Muna–Indonesian version with the purpose of having it checked for mistakes and typos by my counterpart, La Ode Sidu,

and our assistant, Syahruddin. This preliminary version was also presented to the district head and the head of the department of education in 1996. Later I found out that La Ode Sidu had worked hard to get this preliminary version formally acknowledged as a resource book for the schools on Muna. As a result, copies of this version were circulating in the primary and middle schools all over Muna, as I found out during subsequent visits.

Preparing the final Muna–Indonesian version still involved a lot of work. The whole book was reread again, and all mistakes corrected, including those found by the proofreaders. To reduce the number of pages (and hence the cost) for the Indonesian version a smaller font was used and a double-column format. This meant rechecking the whole document after it was printed in order to get rid of overlong white spaces between words. This was accomplished by manually breaking off long words.

Finding an Indonesian publisher for the Muna–Indonesian version was not easy. I contacted a number of publishers on Java, but all required substantial amounts of money to be paid up front ($8,000–$10,000 for 500 copies). I was hoping for a publisher who was willing to take the risk of funding this publication, but of course from a business perspective the market for a vernacular dictionary is very small—essentially only one island in a remote province. Only one publisher was willing to take some risk, but it still required a substantial lump sum, and the final book price was to be rather high for an Indonesian book (around Rp 60,000 or $6.00–$8.00). In the end we had to abandon the idea of working with a commercial publishing house. Instead, on the advice of colleague Charles Grimes, I contacted Artha Wacana Press in Kupang (West-Timor), a young non-commerical academic publisher specializing in eastern Indonesia. They were willing to be the publisher, while a printer in Makassar (Intisari) was prepared to print at a reasonable cost. Still, money had to be paid up front to the printer. My hopes that the provincial Department of Education in Southeast Sulawesi would have money available was unfounded. Fortunately a businessman on Muna who runs a book and photocopy shop was prepared to invest in this project. We agreed on the final price (Rp 25.000 or $2.50–$3.50), and with his investment money for 500 copies I needed only an extra $700 to subsidize the publication. Most of that money came from a SIL hip money project. The Indonesian version was printed in May 2000, shipped to Muna and is now available in the bookshop.

9. What has been most helpful? A number of factors have contributed substantially to the successful completion of this project. In the first place, a limited field period of 18 months provided a strong incentive to treat this project as a manageable undertaking. In my opinion, many dictionary

projects suffer from the fact that there is no fixed end time to the data collection phase; as a consequence the project may become a never–ending enterprise. Also, with such a short fieldwork period, adequate and resourceful planning is an absolute necessity.

In the second place, I was very lucky to have such excellent field collaborators: my official counterpart, La Ode Sidu in Kendari; our trusted village friend, La Ada, whose knowledge of words was truly amazing and whose willingness to help was even more astounding; also our village assistant, Syahruddin, who patiently wrote the many essays on semantic domains.

Thirdly, our move to the Wycliffe Centre in England at the beginning of 1995 put me in an environment where adequate help was available, both in terms of checking English, help with computer issues, and proofreading.

Finally, the promises and encouragements of a sympathetic publisher provided strong incentive to bring the project to fruition. I contacted the KITLV Press at a very early stage of the data collection, and after a provisional printout had been approved by the editorial board, the prospect that the book would actually be printed definitely helped me to keep the end result in view during the whole process.

10. What would I do differently? If I had to do it again (for another language), would I roughly do it the same way? The answer is a modified yes. In addition to the helpful factors that I listed above, I would probably add or change a few matters. Firstly, it would be helpful to have a larger budget so that more local co–workers could be employed and trained, and several computers purchased. Ideally, an office would be set up with several people working concurrently on different parts: some collecting literature, some writing essays, some writing entries, some checking with groups of native speakers, some proofreading, etc. A larger budget would also allow national or foreign specialists to visit the area and identify birds, fish, plants, and trees. If more time, people, and computers were available, the use of concordance lists could also be maximized. With a good e–mail connection on the field, quick advice and answers can also be obtained from others.

Another area that was not really considered—either by myself or by the KITLV Press—is electronic dissemination (see Svensén 1993). Since all the material is on computer, it could easily be disseminated on CD–ROM. This may be especially useful for researchers since the search possibilities are considerably greater than in the printed book. This issue has not been given proper attention for the present project.

However, given the various limitations mentioned earlier, I am very grateful that both the Muna–English and the Muna–Indonesian version have come to fruition.

REFERENCES

Al-Kasimi, Ali M. 1977. Linguistics and bilingual dictionaries. Leiden: Brill.

Bartholomew, Doris and Louise C. Schoenhals. 1983. Bilingual dictionaries for indigenous languages. Mexico: Summer Institute of Linguistics.

van den Berg, René. 1989. A grammar of the Muna language. Dordrecht: Foris.

____(with the cooperation of La Ode Sidu). 1996. Muna–English dictionary. Leiden: KITLV Press.

____ and La Ode Sidu. 2000. Kamus Muna–Indonesia. Kupang: Artha Wacana Press.

Caroll, Vern. 1966. 'Generative elicitation techniques in Polynesian lexicography'. Oceanic Linguistics 5:59–70.

Newell, Leonard E. 1986. Lexicography Notes. Manila: Summer Institute of Linguistics.

____ 1995. Handbook on Lexicography for Philippine and other languages. Manila: Linguistic Society of the Philippines.

Svensén, Bo. 1993. Practical lexicography. Principles and methods of dictionary–making. Oxford: Oxford University Press.

René van den Berg, P. O. Berremah, Darwin, NT 0828, Australia.
E-mail: vandenberg_rene@sil.org

Review Article

Presumptive meanings. By STEPHEN C. LEVINSON.
Cambridge, MA: MIT Press. 2000. 480 pp. Paper $35.00

Reviewed by WILLIAM C. MANN
Language and Translation Dept., JAARS

This review article incorporates substantial parts taken from a review by Dr. Nancy L. Green of University of North Carolina at Greensboro, published in *Computational Linguistics*, 27 (3), September 2001. Used by permission, acknowledged with thanks.

Stephen C. Levinson, (Max Planck Institute for Psycholinguistics, Nijmegen) is a well known contributor to the field of linguistics, especially linguistic pragmatics. He is the author of *Pragmatics* (Levinson 1983) and co–author of a seminal paper on linguistic politeness (Brown and Levinson 1978).

As the title suggests, this linguistic book is on the subject of meanings. This review, with a somewhat wider scope, is also about meanings. In past decades there was an accepted and apparently clear layered series of linguistic abstractions with which to describe a text (written or oral, monologue or interactive). Phonological or orthographic issues were confined to the lowest layer, morphology, the lexicon, and syntax confined to the next layer, and semantics—the meaning of sentences or sentence fragments—was on top. These concerns are no longer seen as neatly layered. Instead, there is rich two way interaction between the various accounts. Texts are not generally seen as independent of their circumstances of creation, the culture, situation, identity, and personal attributes of the speaker or writer. Sentences are no longer seen as fully understandable or analyzable apart from context, which includes the factors above and also the prior language, and which may depend directly on other texts or knowledge about the text recipients.

There have been several significant shifts away from the earlier views. They include a shift from sentences or propositions to some sort of acts as the basic unit of meaning, the recognition of utterances as distinct from sentences, the inclusion of the language producer and recipients as essential to the analysis of meaning, serious attention to varieties and effects of context, expansion away from the common focus on written monologue to include conversation, and the recognition that single utterances or sentences can simultaneously express several varieties of meaning. These trends in the

focus of research represent useful changes for SIL, since SIL translation is focused on multiple authors, texts with cultural and cross cultural issues.

The shift that is most represented in this book is the final one, from THE MEANING to MULTIPLE SIMULTANEOUSLY VALID VARIETIES OF MEANING. We commonly recognize intonation as providing a kind of meaning that supplements the meanings identifiable from a speaker's sequence of words. We commonly recognize pronoun resolution and interpretation of definite determiners as providing meaning which supplements what can be known from the pronouns and determiners themselves. As Levinson shows, other kinds of meaning require more complex investigation.

As indicated below, this work is a direct descendant of work of Grice, which was also the basis for the development of Relevance Theory (Sperber and Wilson 1995).

For SIL, this book has at least three kinds of significance:

- It provides an entry point and a rich overview of the literature of linguistic pragmatics.
- It provides a specific view, perhaps the most detailed and carefully developed in the literature, of a certain kind of implicit meaning.
- It provides a comparison point for Relevance Theory, a comparison which Levinson himself makes.

The first item below is the review of this book by Dr. Nancy Green. Her review is followed by comments on the book from two prominent researchers, Jerrold Sadock and Herbert Clark. Finally, I make additional comments.

FROM NANCY GREEN'S REVIEW

Levinson's book presents a theory of generalized conversational implicature (GCI) and makes a central claim that this theory necessitates a 'new view of the architecture of the theory of meaning' (p. 9). Levinson claims that to account for GCI (and other types of PRESUMPTIVE MEANINGS, or preferred interpretations) it is necessary to distinguish a new level of UTTERANCE–TYPE meaning from sentence–meaning and speaker–meaning: 'This level is to capture the suggestions that the use of an expression of a certain type generally or normally carries, by default' (p. 71). The book belongs to the genre of linguistic argumentation. Expanding upon the Gricean notion of GCI (Grice 1975), the author provides numerous examples of GCI and classifies them into three categories, each category representing a different

licensing heuristic. Then he discusses the implications of the theory: first, for the interface between semantics and pragmatics, and second, for syntactic theory. Throughout the presentation, the author addresses in great detail potential objections and counterarguments from alternate theories of meaning.

According to the author, GCIs are defeasible inferences [treated like ordinary inferences except that they are cancelable in certain principled ways -- ed.] triggered by the speaker's choice of utterance form and lexical items because of three heuristics mutually assumed by speaker and hearer. The heuristics, which can be related to Grice's maxims, are these:

1. The First (Q) Heuristic: 'What isn't said, isn't'. For example, in the context of a blocks world where there are salient oppositions of objects {cones, pyramids, cubes} and colors {red, blue}, from the assertion 'There's a blue pyramid on the red cube,' this heuristic triggers the following inferences: 'There is not a cone on the red cube'; 'There is not a red pyramid on the red cube' (p. 31).

2. The Second (I) Heuristic: 'What is simply described is stereotypically exemplified'. For example, from the assertion 'The blue pyramid is on the red cube,' in the context described above one is licensed to infer 'The pyramid is a stereotypical one ...', 'The pyramid is directly supported by the cube ...', etc. (p. 32).

3. The Third (M) Heuristic: 'What's said in an abnormal way, isn't normal; or Marked message indicates marked situation'. For example, from the assertion 'The blue cuboid block is supported by the red cube', in the context described above one is licensed to infer 'The blue block is not, strictly, a cube', 'The blue block is not directly or centrally or stably supported by the red cube', etc. (p. 33).

In addition, the theory provides a refinement to Gazdar's (1979) projection mechanism; GCIs licensed by heuristic 1 are preferred to those licensed by heuristic 3, which in turn are preferred to those licensed by heuristic 2.

The book argues against the traditional view of the roles of semantics and pragmatics, according to which the output of semantics is the input to pragmatics. Instead it argues for a more complex relationship where GCI can play a role in truth conditions. In this model, two distinct types of semantic processes and two distinct types of pragmatic processes are involved. First, the semantic representation derived from the syntactic structure and lexical items of a sentence may be under–specified. The output of this semantic process is the input to a pragmatic process (Gricean

pragmatics I), in which default, defeasible pragmatic inferences such as GCI may contribute to determining the proposition expressed, e.g. by helping to disambiguate lexical ambiguity, 'generality narrowing' (i.e. narrowing word sense), and determining reference. The output of this process is the input to model-theoretic semantics. After sentence–meaning has been determined by this semantic process, another pragmatic process (Gricean pragmatics II) is responsible for deriving other inferences such as particularized conversational implicatures; this final process yields speaker–meaning.

Certainly, the debate on the role of pragmatics in the linguistic 'architecture' is of significance to computational linguistics. In addition, the book provides a wealth of descriptive information on GCI as well as many pointers to related work in theoretical linguistics. However, unfortunately, the book lacks a computational or formal orientation. For example, much theoretical work would remain to build a computational model of GCI based just upon the information presented in the book. Also, its coverage of potentially relevant work in computational linguistics is not up–to–date; e.g. there is no discussion of recent lexical–pragmatics–oriented approaches, e.g. (Di Eugenio and Webber 1996, Elhadad, McKeown and Robin 1997, Stone and Webber 1998). Despite these limitations, this book will be of interest to language researchers, computationally-oriented or not, with an interest in theories of meaning.

References made by Green

Di Eugenio, Barbara, and Bonnie Webber. 1996. Pragmatic overloading in natural language instructions. International Journal of Expert Systems, 9(2):53–84.

Elhadad, Michael, Kathy McKeown, and Jacques Robin. 1997. Floating Constraints in Lexical Choice. Computational Linguistics 23(2):195–239.

Gazdar, G. 1979. Pragmatics: Implicature, presupposition, and logical form. New York: Academic Press.

Green, Nancy L., and Sandra Carberry. 1999. Interpreting and Generating Indirect Answers. Computational Linguistics 25(3):389–435.

Grice, H. Paul. 1975. Logic and Conversation. In Syntax and Semantics III: Speech Acts, ed. by Peter Cole and Jerry L. Morgan. Academic Press, pages 41–58.

Stone, Matthew, and Bonnie Webber. 1998. Textual economy through close coupling of syntax and semantics. Proceedings of the Ninth International Workshop on Natural Language Generation, Niagara–on–the–Lake, Ontario, Canada, 5–7 August 1998, pp. 178–87.

Nancy Green is an Assistant Professor of Computer Science at the University of North Carolina at Greensboro. Her dissertation research was on a computational model of conversational implicature for interpreting and

BEST COPY AVAILABLE

generating indirect answers (Green and Carberry 1999). Green's address is: Department of Mathematical Sciences, 383 Bryan Building, University of North Carolina at Greensboro, Greensboro, NC 27402–6170; email: nlgreen@uncg.edu

COMMENTS BY SADOCK AND CLARK, WITH PERMISSION FROM MIT PRESS

> *Presumptive Meanings* is a careful and cogent reformulation of linguistic pragmatics based on a sweeping knowledge of the existing literature and a profound understanding of the problem of meaning in natural language. The book presents a significant challenge to both ultra–functionalist views and to the reigning structural theories of the relation between meaning and form. By showing how generalized conversational implicature interacts with both syntactic and semantic description, Levinson has produced the first real advance in the Gricean program in a decade.
>
> —Jerrold Sadock, Glen A. Lloyd Professor of Linguistics, University of Chicago

> *Presumptive Meanings* is a *tour de force* in the study of language and meaning. It takes us beyond what people say, as they rely on the conventional meanings of words and sentences, into the subtle and complex world of what people mean but do not say. Levinson's account of this world is not only scholarly and articulate, but full of insights. In this book Levinson takes up Paul Grice's notion of 'generalized conversational implicatures', argues for its importance, and then offers an elegant refinement. Along the way, he presents a scholarly history of Grice's notion, brings a wide range of phenomena to bear on it, and draws out an impressive array of implications of his view. The result is a remarkable achievement—must reading for any serious student of language and meaning.
>
> —Herbert H. Clark, Department of Psychology, Stanford University

SIL must necessarily continually search out the various linguistic understandings of MEANING, since they are of overwhelming significance to its goals. So the importance of assimilating this major contribution is clear.

The book is but one large item from the now abundant flow of scholarly works in pragmatics, and GCI is but one of the four varieties of meaning, three of them implicit, that were identified in Grice's seminal article. SIL needs a practical grasp of the whole matter, and this book is an advantageous starting point.

As an entry into pragmatics, it has the advantage of putting many works into their technical contexts, but of course it also leaves many parts of pragmatics untouched.

Concerning Relevance Theory (RT), Levinson exhibits numerous contrasts between RT and his account of Generalized Conversational Implicature (GCI). Some of the most comprehensive have to do with situations where RT uses deduction (non–cancelable) and GCI uses implicature, which is cancelable. Other contrasts concern whether effects are entirely context determined and case driven, rather than being subject to more general principles. Levinson's account presents RT with a substantial, even troubling sort of challenge, but it also presents translators and researchers a rich array of ideas with which they may refine their understanding of texts and linguistics. In this book, more than any other single approach, it is RT that is set in contrast. He also by citation links this work to some of his previous publications in which comparisons to RT were made (Levinson 1987; Levinson 1989). He also makes direct comparisons with other systems (Lewis 1979), (Thomason 1987, 1990), (Hobbs 1987).

One of the aspects of quality, for dictionaries, is their usefulness for particular applications. In SIL, using dictionaries in translation, it is important to reflect the lexical side of euphemism, taboo, and the matters that are left implicit—perhaps because they are obvious. Implicit communication complements explicit communication. Dictionaries can usefully indicate what implicit communication commonly accompanies particular words.

According to Levinson, implicit communication occurs in principled ways, and we can project these principles to the very large and small units of text. As an example, section 2.2.3, page 98, discusses a particular principle tied to the lexicon, identifying implicit communication arising from contrast between lexical items. He mentions four distinct cases.

Dictionary builders may wish to make such judgments—about what implicit communication might arise from particular lexical usage—possible for the reader. In any event, there are many sorts of implicit communication that depend directly on lexical items, so a knowledge of these principles will help one to assess what a dictionary can and cannot supply.

I commend this book to your careful attention.

REFERENCES

Brown, P. and Stephen C. Levinson. 1978. Universals in language usage: Politeness phenomena. Questions and Politeness: Strategies in Social Interaction edited by E. Goody. Cambridge: Cambridge University Press 56–289.

Hobbs, J. 1987. Implicature and definite reference. Stanford: CSLI Publications 87–99.

Levinson, Stephen C. 1983. Pragmatics. Cambridge: Cambridge University Press.

____. 1987. Explicature explicated. Behavioral and brain sciences 10:722–723.

____. 1989. Relevance. Journal of Linguistics 21:455–472.

Lewis, D. 1979. Convention: A philosophical study. Cambridge: Harvard University Press.

Sperber, D. and D. Wilson. 1995. Relevance: Communication and cognition. New York: Oxford University Press.

Thomason, R. H. 1987, 1990. Accommodation, meaning and implicature: Interdisciplinary foundations for pragmatics. Intentions in communication ed. by P. Cohen, J. Morgan, and M. E. Pollack. Cambridge: MIT Press 325–364.

William C. Mann, 6739 Cross Creek Estates Rd., Lancaster, SC 29720 USA
E-mail: bill_mann@sil.org

■

European Training Programme (ETP)

We are very pleased to announce that Gene Casad has agreed to come to Horsleys Green to lead a two week course in Cognitive Linguistics. This will be from July 22 to August 2, 2002 and the cost, with full board and lodging included, will be GBP 402.

Gene writes:

> I will aim to acquaint the participants with the main outlines of Ron Langacker's approach to linguistic analysis, description, and explanation known as Cognitive Grammar. The emphasis here will be on conveying a comprehensive view of Ron's approach to Linguistic Semantics. In addition, I will attempt to offer sufficient background in the approach so that participants will be able to read the literature with understanding and to use at least some of the notational devices to make tentative analyses on their own. In addition, I hope to equip them with the needed conceptual tools that they can use to build on the material that will be presented to them in this course.

Please help us by making this course known further afield, and to sign up or to get more information, write to: ETP_UK@sil.org.

Dissertation Abstracts

Proto–Bungku–Tolaki: Reconstruction of its Phonology and Aspects of its Morphosyntax

David Mead
Rice University, Houston, Texas 1998

The Bungku–Tolaki group of languages (Austronesian, Western Malayo-Polynesian) comprises fifteen languages spoken in and around the southeastern peninsula of Sulawesi Island in Indonesia. No written records for these languages existed prior to 1900. To develop a picture of their common ancestor language, Proto–Bungku–Tolaki, I apply the traditional methods of historical and comparative linguistics, as well as more recent understandings regarding the nature of grammatical and semantic change.

The dissertation has two parts. In part one, I reconstruct the sound system of Proto–Bungku–Tolaki, detailing both the innovations which distinguish it from its nearest identified ancestor, Proto–Malayo–Polynesian, as well as the phonological changes which occurred in the various daughter languages, bringing us up to the present day. In the second part focus is on issues of transitivity including the grammaticalization of the preposition **aken* as a valence–changing applicative suffix, clause structure including relative clauses, and verbal inflection. Proto–Bungku–Tolaki is reconstructed as having three construction types which allowed the expression of both an agent and a patent, namely the active, the passive, and the antipassive. Nominative and absolutive pronoun sets served as agreement markers, though the genitive subject marking original to subordinate temporal adverbial clauses has in some languages also made its way into main clauses. Other languages have innovated a second set of nominative pronouns used only in future contexts.

Because there is not as yet a significant body of published material on the Bungku–Tolaki languages, I have made a conscious effort to amply supply this dissertation with the primary data upon which my analyses have been based. Therefore although the present work is of particular relevance to Austronesianists working in the field of historical reconstruction, the data and descriptions alone should make this a valuable reference for anyone interested in the languages of this small corner of the world.

Appendices include five texts with interlinear glossing and free translation, and a compilation of Proto–Bungku–Tolaki lexical reconstructions with supporting evidence.

David Mead, P. O. Box 81439, 6000 Davao City, Philippines. E-mail: David_Mead@sil.org

Confluence, convergence, and contextualization in philosophy and linguistics: The methodologies of Ricoeur and Pike

J. Randolph Radney
The University of Texas at Arlington, 1999

The valuable linguistic insights of Kenneth L. Pike, who developed tagmemic theory, have been formulated independently from the equally valuable work in philosophical hermeneutics done by Paul Ricoeur, who has apparently also been unaware of tagmemic linguistics. Robert E. Longacre has made considerable use of Ricoeur's work in the development of his own version of tagmemics, but this latter version of the theory differs crucially from Pike's. This present work examines areas of agreement in Pike and Ricoeur, particularly as these relate to their views of the nature of humanity and communication and language. The comparison of their work is introductory rather than exhaustive. It seeks to provide a foundation for integration of the two scholars' work into a new, philosophically–based approach to linguistic analysis. This integrated framework will also be shown to provide the means whereby translators may begin to make more systematic use of literary critical theory, semiotics, and this newly developed method of linguistic analysis in their translation work.

The first chapters of the dissertation explore the philosophical foundations of Pike and Ricoeur and develop the framework that begins to integrate their work. A transitional chapter explores various ways in which phenomenology, the philosophical foundation for both Ricoeur's and Pike's work, has been introduced to various disciplines within the humanities. Later chapters show how insights from each of the areas of literary criticism, semiotics, and referential text analysis, may be incorporated into the newly–extended tagmemic theory, especially as it applies to translation. A final chapter briefly explores how the new approach to linguistics provides a context for understanding analyses done in various sub–disciplines of linguistics, such as phonology, survey, and text analysis.

J. Randolph Radney, c/o CanIL, 7600 Glover Rd., Langley, BC V2Y 1Y1 Canada
E-mail: Randy_Radney@sil.org

Minhe Mangghuer: A Mixed Language of the Inner Asian Frontier*

Keith W. Slater
University of California, Santa Barbara, 1998

This dissertation, currently being revised extensively prior to publication, is a grammatical and historical account of the Minhe Mangghuer language. Mangghuer, spoken in China's Qinghai Province, has previously been considered to be a dialect of the Monguor or Tu language, of the Mongolic language family. This research suggests that Mangghuer ought to be considered a separate language.

The dissertation is organized around a systematic synchronic description of Mangghuer phonological, morphological and syntactic structures. These structures are considered from the perspective of a historical, typological, and functional approach to their development and use. Most of the data is drawn from natural discourse, relying primarily on a corpus of folktale narratives. Wherever possible, the Mangghuer system is compared to those of other Mongolic languages. Mangghuer retains many typological features from its Mongolic roots, including SOV basic word order, clause-chaining syntax, and exclusively suffixing, agglutinative morphology, as well as Mongolic core vocabulary. However, Mangghuer has also undergone massive influences from other languages, and the Qinghai–Gansu region in which it is spoken is shown to exhibit many characteristics of a linguistic area. The historical discussion highlights some specific results of language contact on Mangghuer, including massive lexical borrowing from Chinese, as well as the primarily Sinitic phonological system and the Bodic evidential system which Mangghuer has acquired.

Mangghuer cannot be considered simply a Mongolic language in the normal genetic sense. It has features of both a Mongolic language and a mixed language. A grammaticalization perspective yields enlightening accounts of some Mangghuer morphosyntactic developments; it is also suggested that some Mangghuer grammaticalization processes may have been motivated by contact with other languages. The syntactic behavior of Mangghuer nonfinite clauses is shown to challenge some standard analytical terms in the

---

* A substantially revised version of this dissertation is scheduled to be published in 2001 or 2002 by Curzon Press under the title *'Minhe Mangghuer: A Mongolic language of China's Qinghai-Gansu Sprachbund'*.

typology of clause–combining syntax. The distinctions of subordination vs. cosubordination, and adverbial clauses vs. clause chaining are found to be difficult to apply to Mangghuer constructions. The Mangghuer evidence suggests that these categories cannot be held to be rigidly distinct.

Keith Slater, P. O. Box 307, 50000 Chiang Mai, Thailand. E-mail: keith_slater@sil.org

Reviews

Language and the lexicon: An introduction. By DAVID SINGLETON.
New York: Oxford University Press. 2000.
244 pp. Cloth $65.00, paper $19.95.

Reviewed by ROBERT D. EATON
SIL—South Asia Group

Language and the Lexicon is a textbook for a course in lexicography or for anyone involved in creating dictionaries in a form very accessible to beginning linguists. Throughout the book he defines a broad spectrum of terminology and examines the relationship between the lexicon and nearly every other area of linguistic study: from the lower levels of linguistic inquiry, such as phonetics, phonology, and orthography, to the higher levels of syntax, semantics, and sociolinguistics. Each chapter has a helpful summary, an extensive annotated bibliography for further reference, focusing questions, and topics for discussion.

In the introductory material, S develops a working definition of a WORD. He then sets the stage with the lexicon as the main character and previews how it interrelates with the different areas of linguistic study elaborated on in subsequent chapters.

In examining the relationship between the lexicon and syntax, S discusses a variety of theoretical perspectives: computational linguistics (where the lexicon literally drives the computational analysis), Valency Grammar (where verbs and their argument counts are maintained in the lexicon), Lexical–Functional Grammar (where semantic roles and grammatical relations associated with verbs are maintained in the lexicon), and finally, the Chomskyan perspective (where the lexicon is called upon more and more to explain certain idiosyncratic phenomena). He argues persuasively that the growing consensus among these different perspectives is that the syntactic shape of a sentence is largely a function of the lexical elements with which it is made.

The chapter on lexical partnerships is especially helpful for those involved in second language acquisition and translation. S points out how certain lexical choices are dependent on other lexical choices. He gives the example phrase fragment, *at this moment in* ___, which can only easily take *time* as the missing constituent. Besides idiomatic expressions of this type, he argues that even simple compounds often should have independent existence in the lexicon: *blackboard*, for example, is a technical term that means more than

the sum of its constituent parts. Furthermore, the meaning of some COLLOCATIONS (as he calls these partnerships) cannot be straightforwardly computed from the pieces of which they are made: A *heavy smoker* (as opposed to a *heavy vehicle*) has nothing to do with weight and a *criminal lawyer* in most contexts is taken to mean something other than 'law–breaking attorney'.

In the chapters dealing with diglossia, the relationship between the lexicon and sociolinguistics, S points out that different language varieties are often used in different situations (called domains), with the high prestige variety used in formal situations, such as education and literature, and the low prestige variety used in more informal and intimate contexts, such as when talking with friends and family.

The chapter on the relationship between the lexicon and semantics has some helpful definitions that shed light on the structuralist, componential analysis, and cognitive approaches to meaning. Relating to the structuralist perspective, he explains the difference between polysemy (multiple related senses of the same word; e.g. *he has a small MOUTH* vs. *the MOUTH of the river*), homonymy (same form but otherwise unrelated words, e.g. *I deposited the money in the BANK* vs. *I'll meet you at the BANK of the river*), homophones (homonyms which are pronounced the same, but spelled differently; e.g. *MEAT, MEET*), and homographs (homonyms which are spelled the same, but pronounced differently e.g. *row* 'propel a boat using oars' vs. *row* 'quarrel'). He defines other terms that have to do with the relationship between pairs of words, such as hyponymy (the relation between more specific and less specific terms; e.g. a cocker spaniel is a spaniel is a dog is a mammal is an animal), metonymy (part–whole relationships; e.g. a finger is part of a hand, a hand is part of an arm), complementarity (where the assertion of one word is the denial of the other; e.g. *alive–dead* and *pass–fail*), and polar antonymy (like complementarity, but with more possibility of gradation; e.g. *rich–poor*).

In summary, S has done an effective job of presenting a broad spectrum of information in a manner that even novice students can understand. By looking at the intersection of the lexicon with the different areas of linguistic study, the reader gains an overview of most of the field of linguistics. As an introductory textbook for a course in lexicography or as a concise, yet readable introduction to linguistic theory, *Language and the Lexicon* is an excellent offering.

Robert D. Eaton, 4495 W. 225th St., Fairview Park, OH 44156 E-mail: Bob_Eaton@sall.com

BEST COPY AVAILABLE

Modern lexicography: An introduction. By HENRI BÉJOINT. New York: Oxford University Press. 276 pages. Paper $27.95

Reviewed by GILLES GRAVELLE
SIL—Indonesia Branch

Modern Lexicography: An Introduction is not necessarily a text book on how to do lexicography. It is actually a metalexicography book. It is literature about dictionaries, particularly English dictionaries from a French meta–lexicographer's point of view. Focal questions that the author asks are, why are English dictionaries the way they are and are they adapted for our times?

Chapters 1-3 provide interesting and important information regarding the history and development of English dictionaries. The author describes the social and political motivation behind the making of dictionaries. He then raises the questions: Are dictionaries today objective? How strongly do they reflect political and ideological biases? Are they representative of a language? Can they really be representative of a language? Are they prescriptive or descriptive? Should they draw from a text corpus or from the best usage usually found in literature? He stresses the importance of understanding the social and political influences on lexicography throughout history in order to address, later in the book, what the current problems of dictionary–making are, and how lexicographers might improve on them.

Chapters four and five, on general–purpose dictionaries and society, begin to address some practical questions that may be helpful to field linguists who plan to work, or are currently working, on vernacular language dictionaries. The focal question in these chapters is, what is the purpose of a dictionary? The author addresses both philosophical and practical issues in trying to provide some answers. The question seems straightforward until he begins to survey the opinions of a number of lexicographers—particularly those since the 1960s. According to some, it is not just a tool to define words or aid in spelling or pronunciation. It also is a record of the language. It helps to improve communication. It strengthens the language. It is an aid for foreign–language learning. It is an ideological weapon. It is a repository of linguistic truth. It reflects the values of society.

Evidently determining the purpose and structure of a dictionary has been highly subjective and rather prescriptive over the years. The author suggests that anyone attempting to produce a dictionary should first study the population who will use the dictionary to determine what the content and layout should be. 'What a novel idea!' one might say. But until the 1960s, very little research was conducted to find out how useful a dictionary would

BEST COPY AVAILABLE

be to the average user. Chapter five deals extensively with various kinds of testing conducted on different age groups and levels of society to determine how the users actually used the dictionaries, and what they wanted most in a dictionary.

In chapters six and seven the author addresses the linguistic traditions of lexicography. These chapters are particularly interesting and maybe even liberating because one learns, according to the author, that there are no real linguistic traditions in lexicography. Historically, lexicographers did what they did because they had no theories on which to draw—only tradition. Hence the work of lexicographers has been driven more by personal preference and the commercial interest of their publishers more than anything else. Lexicographers determined that the best dictionaries should contain the most important words. The most important words are said to be the most useful. The most useful words are also the most common. But surveys of dictionary users revealed that the most common words are the least consulted. The less common words are the ones that users need more but those are the ones often left out. To publishers, the more entries there are the greater the chance of commercial success. Hence many dictionaries have been produced to include as many words as possible. But according to the author's research, that has not been very useful for the users. Still, the user wants to know that THEIR dictionary has ALL the words in their language. So publishers claim that their dictionaries are exhaustive. The author challenges that notion by pointing out that it is impossible to be truly exhaustive because language does not form a closed set.

Finally, the author addresses the more difficult 'linguistic' issues of dictionary making, namely the problem of defining words, defining what a unit is, and whether multi–word units can or should be included in the dictionary. This section of the book is where my interest became more keenly aroused because it addresses the practical problems that field linguists face on a regular basis in developing lexicons of non–English, indeed non–Indo–European, vernacular languages. Determining what a semantic unit is in Meyah, a Papuan language of Irian Jaya, Indonesia, is difficult when operating on the principles or traditions of English language lexicography. According to the author, in English lexicography there is no linguistic approach to dealing with phrasal units such as idioms, metaphors, and various kinds of collocations. Tradition still dictates that entries should only be single words without a context. But in Meyah all psychological and physical states, such as 'angry', 'weak', and 'sick' are only expressed as body–part metaphors. They are phrasal units. Their meaning is not the sum of each constituent. Many alienable nouns, especially introduced objects, are compounds. Many events are expressed by juxtaposed verbs roots or

verbs plus adjuncts. Not to include multiple word units in the dictionary would be to exclude a large number of verbal and nominal senses that cannot be expressed any other way.

How should phrasal units be organized in the Meyah dictionary? In Meyah, person and number of subject is unmarked (zero morpheme) for third person singular. As a consequence, dictionary entries are alphabetized based on the 3SG forms. Yet a large majority of Meyah psychological state expressions begin with the word *odou* 'liver', as in *odou okum* 'liver heavy' (lit. angry). Following traditional lexicography practices could result in listing all *odou*-phrases as subentries under the noun *odou* 'liver'. This would be miss–leading because it does not differentiate the noun 'liver' from the many verbal senses that words used in combination with *odou* communicate. Also, they are not synonyms.

The book also caused me to ponder more deeply what the purpose of a Meyah dictionary should be and why I might want to spend so many tedious hours working on it. Considering that the Meyah language is a newly written language the dictionary provides a snapshot of its lexicon at a certain point in time. It provides research material for comparative studies. It validates the language in the eyes of the native speakers. However, one wonders if the Meyah people really need it? After 16 years I have only managed to record about 2,300 entries based on a text corpus of mixed genre.[1] It seems that a Meyah native speaker would already know the definitions and usage of so few words, not to mention the possible collocational usages. It could function as a spelling guide, but seeking to establish 'correct' spelling or pronunciation is a tricky undertaking considering that speakers of the various Meyah regional dialects all consider their own spelling/pronunciation to be the correct one? So they do not need a dictionary in the same sense that English speakers might need one. The dictionary could serve as a record representing the day–to–day vocabulary of Meyah. However, the author of *Modern Lexicography* challenges whether any dictionary can truly be representative since it is usually based on the work (and opinion) of only a few. As a record it could contain far more animate and inanimate nouns as well as proper nouns, thus be more

---

[1] For more information on why Meyah, like many Papuan languages, has such an apparent shortage of vocabulary items, see Andrew Pawley. 1993. A language which defies description by ordinary Means. In The Role of Theory in Language Description ed. by William A. Foley, pp. 87–129. Trends in Linguistics. Studies and Monographs 69. Germany: Mouton de Gruyter. 1993

encyclopedic, as many American English dictionaries are. But that begs the question, do the Meyah people need a record of all the nouns in their environment? Would they use the Meyah dictionary to find out what a tree frog is? A quick survey of dictionaries produced by SIL fieldworkers shows that the vast majority are bilingual. Therefore it appears that bilingual education is the primary reason why SIL produces dictionaries.

Modern Lexicography was very useful in helping me understand how much 'tradition' and how very little linguistic theory has influenced English lexicography. It aided my understanding of how I sometimes view the Meyah lexicon from a traditional English lexicography perspective, a disadvantage in my opinion. The book addresses many important issues that traditional English lexicography has yet to come to terms with. Field linguists developing lexicons of vernacular languages have a great advantage. Usually, there are no local lexicography traditions that might overshadow attempts to make truly useful dictionaries, and this book provides much food for thought as we consider how to go about doing that. It is also a reminder for the need to heed lessons about prescriptivism and to work to discover what the users of a dictionary really want or need.

For a more practical step–by–step guide for doing lexicography, see Leonard E. Newell (1994) and Coward and Grimes (1995).

References

Coward, David F. and Charles E. Grimes. 1995. *Making dictionaries: A guide to lexicography and the Multi–Dictionary Formatter (Version 1.0)*. Waxhaw: Summer Institute of Linguistics. ix, 234 p.

Newell, Leonard E. 1994. *Handbook on lexicography for Philippine and other languages*. Linguistic Society of the Philippines, Special Monograph Issue, 36. Manila: Linguistic Society of the Philippines. x, 368 p.

Gilles Gravelle, 26229 222nd Pl. SE, Maple Valley, WA 98038. E-mail: gilles_gravelle@sil.org

Mentor courses: A resource book for trainer–trainers. By Angi Malderez and Caroline Bodoczky. New York: Cambridge University Press. 1999. 220 pp. Hardback $59.95, paper $22.95.

Reviewed by Lou Hohulin
SIL—International Consultant

This book, according to the authors, is meant to be a resource book for trainers. It is based on the development of a new and innovative English

language teacher training program at the Centre for English Teacher Training at Eotvos Lorand University in Budapest, Hungary.

The introduction to the book is excellent. The authors explain the purpose of the book, their audience, a description of a mentor, the organization of the chapters, and the organization of the activities. Although the book is intended for trainers of teachers of English, there are many principles that might apply to any mentoring situation.

Of the 220 pages of the book, 108 pages are devoted to course activities. Mentor course principles and practice are covered in 22 pages; course procedures are covered in 5 and there are 40 pages of photo-copiable resources and 10 appendices with different feed–back and evaluation forms. It is obvious from the allocation of pages that the designers of this course consider application far more important than theory. That fits with the concept of mentoring. Good mentoring is not so much transfer of theoretical information as it is modeling and helping in the actual application of theory in practice.

Part I. Mentor course principles and practice. Part I has two chapters: Basic concepts and Course procedures. These chapters set the mentoring course within a framework of what the authors call a 'constructionist view of learning'. This view of learning sets up challenges for the learner to organize their own behavior and experiences to produce changes in their own existing 'constructs' of knowledge, experience, and beliefs to those that they personally value as 'good teaching models'. Closely related to the constructionist view of learning are the course processes advocated by the authors. 'Social constructivism' is the basis for these processes; this framework takes the context of learning into consideration—the learners themselves, the classroom culture, and even the broader context of the educational system and society. The theoretical foundation of 'social constructivism' appears to be the view that knowledge is jointly and socially created through interaction with others. As might be expected, that viewpoint influences all of the classroom activities presented in Part II.

Part II. In–session activities. For each of the activities the authors have a carefully designed, easy–to–use framework that consists of a general introduction, an aim, suggested position in a sequence of activities, materials needed, time to be allotted for completion of the activity, classroom organization, a clear procedure, author's comments, and participants' reactions (excerpted from evaluations of previous uses of the activity).

Chapter 3, Lead–ins, describes 21 activities that can be used to start the group process, a new session, or a new mentoring topic. Chapter 4 describes

10 activities for developing the mentor skills of observing lessons. According to the authors, one of the most useful things that a mentor can do is to watch a mentee teaching. They say that observing appropriately is a difficult skill to learn and requires practice. Chapter 5 is titled Challenging Appropriately and describes and discusses 16 activities that will encourage participants to consider their role in relationship to aims, objectives, and possible dangers. It also covers the development of the interpersonal skills needed for counseling and challenging appropriately.

Chapter 6 covers role–play activities and the tasks associated with them. The authors consider these activities as the most valuable in mentor courses. They say that these activities (p. 110):

> ... provide time for whole skill practice, a safe space for experimentation, an opportunity for participants to discover their natural or preferred mentor–style, and a range of situations in which to develop the ability to empathize with the mentee.

In all, there are ten role–play activities and all but one have photo–copiable resources in a separate section of the book. Chapter 7 focuses on 'assessing teaching' and discusses the dangers of outsider evaluation and of using standardized assessment tools. Several activities are described that explore the problems related to assessment. The authors stress the importance of developing the student–teacher's ability to self–evaluate (p. 122):

> ... clearly it is the mentor's duty to ensure that mentees not only reach the standards required by their particular context, but that they too have internalised a set of standards by which they can self–evaluate in the future, and from which they can develop their own personal set of criteria.

Part III. Projects and assignments. Part III has 3 chapters: Chapter 8 describes how to make a participant's own teaching the focus of observation and to practice 'reflective behavior'. The activities are classified on the types of tasks: self–observation, peer–observation, and drawing conclusions. Chapter 9, 'Reading tasks' gives ideas for integrating existing outsider theories into participants' personal understandings of mentoring. Chapter 10 contains a description of types of writing tasks and activities. One is the choice of writing topics that relate to the professional learning cycle and the other is the writing of development reports.

Finally, the concluding section of the book describes some activities for closure of the course and disbanding the participant group.

This is an excellent reference book for anyone who might teach a course on mentoring student–teachers or for anyone who is assigned to 'supervise' or

mentor student-teachers. It would take a good deal of thought and study to adapt the principles and activities to other mentoring situations, though it could probably be done.

Lou Hohulin, 7500 W. Camp Wisdom Rd., Dallas, TX 75236.
E-mail: Lou_Hohulin@sil.org

Wilhelm von Humboldt: On Language: On the diversity of human language construction and its influence on the mental development of the human species. Ed. by MICHAEL LOSONSKY. Cambridge Texts in the History of Philosophy. Cambridge: Cambridge University Press. 1999. 296 pp. Hardback $59.95, paper $19.95.

Reviewed by ALAN S. KAYE
California State University, Fullerton

Baron Karl Wilhelm von Humboldt was born in Potsdam, Germany in 1767 and died in Berlin (then known as Tegel) in 1835. He is well known among linguists as an early generativist or, at least, a Cartesian anticipator of transformational-generative grammar, although Hans Aarslef rejects the Cartesianism of Humboldt (Aarslef 1982). Chomsky (2000:6) asserts that Humboldt was, in fact, one of the founders of 'the first cognitive revolution'. I believe most modern-day linguists know Humboldt's name through the writings of Noam Chomsky. Humboldt's famous maxim, 'Language makes infinite use of finite means' (for this quotation see, most recently, Chomsky [ibid.]), has been championed by many a teacher of linguistics. In addition to being a linguist, Humboldt was a statesman: working in 1790–91 as a law clerk in the Prussian Supreme Court (having studied law at the University of Frankfurt/Oder in 1787–88), representing Prussia in the Vatican (1801–08), and serving as the Prussian ambassador in Vienna and London (1810–19). His most visible legacy today is the university named after him, Humboldt Universität in East Berlin, which was founded in 1809 as Friedrich Wilhelm Universität. It should not be forgotten that Humboldt was very well trained at the University of Göttingen in classical philology, even publishing translations of Xenophon and Plato, and, as for many others of his era, there was little separating philology from

linguistics.[1] This fact held true in many linguistic circles through many of the decades of the 20th century.

Most of the present volume is the elegant English translation by Peter Heath (which was first published in 1988 by Cambridge University Press) of Humboldt's *Über die Verschiedenheit des Menschlichen Sprachbaues und ihren Einfluss auf die geistige Entwicklung des Menschengeschlecht* (in English: *On the Diversity of Human Language Construction and its Influence on the Mental Development of the Human Species*). This text was intended to be the introduction to Humboldt's magnum opus, *Über die Kawi Sprache auf der Insel von Java*. The first, second, and third volumes of that work were published in 1836, 1838, and 1840 respectively (all by the Royal Prussian Academy of Science).

The preface to the translation (*On the Diversity* ...) is another translation, viz., that of his brother's (Alexander von Humboldt, the famous scientist and explorer) remarks of March 1836, introducing the posthumous publication of Wilhelm's research (pp. 3–6). The following passage shows the revered esteem Wilhelm received from the fellow scholar that was his brother (p. 4):

> If it was granted to my lamented brother, by the power of his intelligence no less than by the strength of his will, by the favour of outward circumstances, and by studies which his frequent changes of residence and public mode of life were unable to interrupt, to penetrate deeper into the structure of a larger number of languages than have probably yet been encompassed by any single mind, we may doubly rejoice at finding the last, I might well say the highest, results of these researches, which cover the whole domain of language, unfolded in the Introduction to the present work.

The editor's introduction contains a fine summary of Humboldt's career (pp. vii-xxxiv). He spent years studying numerous exotic languages. From 1824–26, he concentrated on Native American languages. From 1827–29, he focused on Sanskrit and the languages of the South Pacific, and from 1829 until his death, he concentrated on 'Malayan' languages (p. x).[2] Surely the enormous diversity of languages with which he came into contact influenced his conception of 'linguistic relativity', and most scholars of the Sapir–Whorf Hypothesis trace this legacy back to his writing: 'Just as

---

[1]As one reads through this volume, one notices Humboldt's intimate familiarity with the details of many languages from his detailed reading of the standard grammars of his day, e.g. the Hebrew treatises by G. H. A. von Ewald (1836, 1870) (fn. 3, p. 79).

[2]The Malayan languages today would be known as Indonesian, a subgroup of the phylum called Malayo–Polynesian or Austronesian.

BEST COPY AVAILABLE

without language no concept is possible, so also [without language] there can be no object for the mind' (p. xvii), and 'since language is involved in structuring human cognition, and languages are diverse in their structure, there resides in every language a characteristic world–view (Weltansicht)'.[3] The Humboldtian view states that there is no thinking which can occur without words. Of course, one of the presuppositions of the Sapir–Whorf Hypothesis, however, is that thinking means linguistic thinking.

The editor refers to sharp criticisms of Chomsky by scholars influential in Germany (p. xxx and passim). These critics point out that Chomsky and Humboldt differ about what is infinite in language structure. Chomsky thinks of language as an infinite number of possible grammatical sentences, but for Humboldt, they claim, it is not infinite sentences but rather infinite thoughts which are key. The editor dismisses the critiques of Chomsky's reading of Humboldt by saying he does not 'think that these features are incompatible with the generative model in linguistics, and, more broadly, in psychology' (ibid.). It stands to reason, in my view, that if every language consists of an infinite number of possible sentences, then those sentences are capable of expressing an infinite number of thoughts. Thus I agree with the editor's conclusions. However, I do not agree with his pronouncement that for Chomsky, the basic unit of language is the sentence, whereas for Humboldt, it is the word (p. xxxii). Although, to be sure, Humboldt wrote more about words and morphology, even the editor asserts that 'if pressed, Humboldt would point to sentences as the linguistic items that have the task of confronting and constituting the boundless domain of thoughts human beings can think' (ibid.).

Even after the 166 years which have elapsed since Humboldt's death, his writings continue to generate new works, including a two–volume biography (Sweet 1978–80), the exhaustive publication of seventeen volumes of his collected writings (Humboldt 1903–36), a five–volume collected writings (Humboldt 1960–81), and a treatise on the philosophical foundations of the linguistic theory underlying his work (Manchester 1985). One can justifiably conclude from this plethora of research activity over the last decades that sometimes new is not necessarily better. Occasionally, it is old which is better. Today's linguist will be much better off examining the work

[3]As is well known, the concept 'world-view', or *Weltanschauung*, has received voluminous attention by many scholars, and this idea has continued to fascinate the psychological, anthropological, and linguistic worlds.

of this European giant,[4] and then, too, there is Neo–Humboldtian linguistics which is surely an important field of inquiry.[5]

REFERENCES

Aarslef, Hans. 1982. From Locke to Saussure. Minneapolis: University of Minnesota Press.

Chomsky, Noam. 1965. Aspects of the theory of syntax. Cambridge, Mass.: MIT Press.

____. 2000. New horizons in the study of language and mind. Cambridge: Cambridge University Press.

Humboldt, Wilhelm von. 1903–36. Gesammelte Schriften. 17 volumes. Berlin: Preussische Akademie der Wissenschaften. Reprinted, Berlin: Mouton de Gruyter, 1967–68.

____. 1960–81. Werke. 5 volumes. Darmstadt: Wissenschaftliche Buchgesellschaft.

Manchester, Martin L. 1985. The philosophical foundations of Humboldt's linguistic doctrines. Amsterdam: John Benjamins.

Sweet, Paul R. 1978–80. Wilhelm von Humboldt: A biography. 2 volumes. Columbus: Ohio State University Press.

Von Ewald, G. H. A. 1836. A grammar of the Hebrew language of the Old Testament. Tr. John Nicholson. London: Whittaker.

____. 1870. Ausführliches Lehrbuch der hebräischen Sprache, 3rd ed., Göttingen, 1870.

Alan S. Kaye, California State University, Fullerton, CA 92634–9480.
E–mail: akaye@Exchange.FULLERTON.EDU

Presuppositions and anaphora. By EMIEL KRAHMER. New York: Cambridge University Press (Stanford: CSLI). 1998. 255 pp. Hardback $59.95, paper $22.95.

Reviewed by CHRISTOPH UNGER
SIL—Eurasia Area

This book revisits presuppositional phenomena in the light of modern formal semantic approaches which look at the meaning of sentence sequences

---

[4]He kept very good European company, and was a personal and intimate friend of Johann Christoph Friedrich von Schiller (1759–1805) and Johann Wolfgang Goethe (1749–1832).

[5]Space prohibits our entering into details in this regard, unfortunately.

BEST COPY AVAILABLE

(discourses) rather than on the meaning of sentences alone, so–called dynamic approaches to semantics.

Dynamic semantic systems have been developed primarily in the light of cross–sentential anaphors. Sentences containing anaphors such as *She kissed it* (Krahmer's example (1), p. 1) have widely different meanings depending on how the anaphors are interpreted in context. It appears that the meaning of a sentence needs to be specified not in a static way referring to its logical properties in isolation, but in a dynamic way that incorporates effects of preceding sentences in a formal way.

The shift from static semantic systems to dynamic ones which has taken place since about the early 1980s has lead semanticists to re–assess old problems in the light of the new dynamic approaches. The treatment of presuppositions is one such problem. The most notorious problem concerning presuppositions is the projection problem. While in some cases presuppositions survive under logical connectors, in others they don't:

(1) If the king of France needs a new chambermaid, then his wife will keep a close watch on the course of things.

Presupposes: there is a king of France—example (15a) on p. 5.

(2) If there actually is a king of France, then the king of France needs a new chambermaid.

Does not presuppose: there is a king of France—example (15b) on p. 5.

It has been argued in the tradition of Frege and Strawson that presuppositions are best treated in logical systems which incorporate a truth–value gap (or third truth–value, depending on one's perspective) saying that when the presupposition turns out to be false (e.g. when there is no king of France), the sentence is neither true nor false rather than false (such systems are called PARTIAL LOGICS by Krahmer). However, static partial systems lack the flexibility required to account for the projection problem. Therefore partial logic–based approaches to presuppositions were largely abandoned in the 1970s and 1980s.

Krahmer's main purpose is to rehabilitate partial–logic based semantics. The overall structure of his argument is this: the dynamic view of formal semantics leads to an improvement in the theory of presuppositions by being able to relate presuppositions to anaphora in van der Sandt's presuppostion–as–anaphor theory (van der Sandt 1989, 1992, van der Sandt and Geurts 1991) based on discourse representation theory (DRT). This DRT based approach can be further improved by techniques developed in partial logics and by having access to partial logic–based semantic interpretation in the

discourse structure–building algorithm. To make this work, Krahmer argues in particular that there is a version of partial logic which has the flexibility required to deal with the projection problem of presuppositions independently of the dynamic part of the resulting formalism.

This overall argument is indeed an ingenious one. It is one that is run not so much by introducing new approaches but by assessing existing approaches, extending and sharpening them in some details, and putting them together in a common perspective. This is a most welcome achievement in a field that seems to be rife with a proliferation of different semantic and logical formalisms. Consequently, as a by–product, the reader is given a good bibliographical overview of the field.

Whether the argument is forceful enough to convince skeptics of partial logic–based approaches to presuppositions, such as the author of this review, is a different question. For reasons of space only a few general objections are outlined. First, Krahmer gains the flexibility of partial logics by introducing an assertion operator. Thus, in effect (3a) is analysed as (3b), which obviously does not presuppose that there is a king of France:

(3) a. The king of France is not bald, there is no king of France.
b. The speaker does not assert that the king of France is bald, there is no king of France.

Krahmer shows rightly that with this operator one does not have to postulate an otherwise unmotivated introduction of a second negation operator. However, this result is bought with the introduction of a different operator and a sophisticated machinery to constrain its application. This apparently misses the force of the traditional argument, which was that partial logics can cope with the data only by making the formal system gratuitously complex. Krahmer's assertion operator mechanism seems to be subject to the same criticism. Hence the force of Krahmer's argument is at least greatly reduced.

A second criticism is connected with a technique used in constructing a semantics for partial logics: that of characterizing the meaning of a sentence by a disjunction of its truth– and falsity–conditions respectively. This technique seems to predict rather unintuitive presuppositions. Thus, (4) is predicted to presuppose (5) (Krahmer's examples (12) on page 138–139):

(4) Somebody stopped dating Mary.
(5) Either somebody has been dating Mary before and now no longer dates her, or everybody dated Mary before and still dates her at the moment.

But this is not an intuitive description of what is presupposed here. Note that (5) is logically equivalent with the simpler (6):

(6) Somebody has been dating Mary before and now no longer dates her.

Turning to cognition, the mind tends to allocate processing resources to that information which seems to be most relevant (in a technical sense) (Sperber and Wilson 1995; 1996). Since (5) is less costly to process (since it is less logically complex), it is more relevant than (6). Hence (6) stands little chance to be mentally processed. The upshot of this discussion is that the technique Krahmer uses in analyzing presuppositions can predict irrelevant outputs which the mind is not disposed to process. This casts doubts on the adequacy of semantic approaches making use of these techniques.

This book is written for specialist readers with the necessary background in formal semantics rather than for field workers in descriptive linguistics. However, the reader with the necessary background in model–theoretic semantics—whether agreeing with Krahmer's position or not—will find a wealth of ideas which may even influence his or her descriptive fieldwork by studying definite descriptions, anaphora, and presuppositions in the language under investigation with ideas from this book in mind.

REFERENCES

Sperber, D., and D. Wilson. 1995. Relevance. 2nd edition. Oxford: Blackwell. (1st edition 1986).

____ 1996. Fodor's frame problem and relevance theory. Behavioral and Brain Sciences 19.530–532.

Van der Sandt, R. 1989. Presupposition and discourse structure. Semantics and Contextual Expression, R. Bartsch, J. van Benthem, and P. van Emde Boas, eds. Dordrecht: Foris Publications.

____. 1992. Presupposition projection as anaphora resolution. Journal of Semantics. 9.333–377.

____ and B. Geurts. 1991. 'Presupposition, anaphora and lexical content'. Text Understanding in LILOG, O. Herzog and C.-R. Rollinger, ed. Berlin: Springer Verlag.

Christoph Unger, Siegenweg 31, 57299 Burbach, Germany.
E-mail: Christoph-Kuelvi_Unger@sil.org

Books Available for Review

The following books are available for review by our readers. If you wish to do a book review for publication in *Notes on Linguistics*, contact the editor and the book will be shipped to you along with instructions for submitting the review. When you submit a review, the book is yours to keep. Contact:

Mike Cahill, Editor, Notes on Linguistics
7500 W. Camp Wisdom Road, Dallas, TX 75236
Internet e-mail: Eugenecc_Loos@sil.org. (Editorial Assistant)

Ackerman, Farrell & John Moore. 2001. Proto–properties and grammatical encoding: A correspondence theory of argument selection. Chicago: The University of Chicago Press (Distributed for CSLI Publications). 197 pp. Cloth $55.95, paper $18.95.

Adger, David; Susan Pintzuk; Bernadette Plunkett; & George Tsoulas, eds. 1999. Specifiers: Minimalist approaches. New York: Oxford University Press. 344 pp. Paperback $45.00.

Benmamoun, Elabbas. 2000. The feature structure of functional categories. New York: Oxford University Press. 180 pp. Cloth $45.00, paper $19.95.

Brown, Sue. 1999. The syntax of negation in Russian: A Minimalist approach. New York: Cambridge University Press. 154 pp. Hardback $49.95, paper $18.95.

Bucholtz, Mary; A. C. Liang; & Laurel A. Sutton, eds. 1999. Reinventing identities: The gendered self in discourse. 431 pp. Cloth $65.00, paper $35.00

Chapelle, Carol A. 2001. Computer applications in second language acquisition: Foundations for teaching, testing and research. New York: Cambridge University Press. 215 pp. Hardback $59.95, paperback $21.95.

Choi, Hye-Won. 1999. Optimizing structure in context: Scrambling and information structure. New York: Cambridge University Press. 223 pp. Hardback $59.95, paper $22.95.

Costa, João. 2000. Portuguese syntax: New comparative studies. New York: Oxford University Press. 304 pp. Cloth $65.00, paper $45.00

Cowie, A.P., ed. 2001. Phraseology: Theory, analysis, and applications. New York: Oxford University Press. 258 pp. Cloth $24.95.

Crain, Stephen & Rosalind Thornton. 2000. Investigations in universal grammar: A guide to experiments on the acquisition of syntax and semantics. Cambridge, MA: The MIT Press. 360 pp. $27.00

Dahllöf, Mats. 1999. Three papers on computational syntax and semantics. Reports from Uppsala University, Dept. of Linguistics. 123 pp.

Dalrymple, Mary, ed. 1999. Semantics and syntax in lexical functional grammar: The resource logic approach. Cambridge, MA: The MIT Press. 399 pp. Hard cover $40.00

Davidson, Donald. 2001. Inquiries into truth and interpretation. New York: Oxford University Press (Clarendon Press). 296 pp. Paper $21.95.

de Boer, Bart. 2001. The origins of vowel systems. New York: Oxford University Press. 168 pp. Hardcover $55.00.

Dekkers, Joost; Frank van der Leeuw; & Jeroen van de Weijer. 2000. Optimality theory: Phonology, syntax, and acquisition. 635 pp. Paper $45.00.

Deuchar, Margaret & Suzanne Quay. 2001. Bilingual acquisition: Theoretical implications of a case study. New York: Oxford University Press. 163 pp. Paper $18.95.

Görlach, Manfred, ed. 2001. A dictionary of European Anglicisms: A usage dictionary of Anglicisms in sixteen European languages. New York: Oxford University Press. 352 pp. Hardcover $99.00.

Gramley, Stephan. 2001. The vocabulary of world English. New York: Oxford University Press. 323 pp. Paper $21.95.

Haspelmath, Martin. 2001. Indefinite pronouns. New York: Oxford University Press. 364 pp. Paper $29.95.

Kenstowicz, Michael, ed. 2001. Ken Hale: A life in language. Cambridge, MA: The MIT Press. 482 pp. Paper $30.

King, Jeffrey C. 2001. Complex demonstratives: A quanificational account. Cambridge: The MIT Press. 207 pp. Cloth $45.00, paper $18.00.

Kuteva, Tania. 2001. Auxiliation: An enquiry into the nature of grammaticalization. New York: Oxford University Press. 209 pp. Hardcover $70.00.

Lycan, William G. 2001. Real conditionals. New York: Oxford University Press. 223 pp. Hardcover $35.00

Maffi, Luisa, ed. 2001. On biocultural diversity: Linking language, knowledge, and the environment. 578 pp. Cloth $65.00, paper $34.95.

Matthews, Peter. 2001. A short history of structural linguistics. New York: Cambridge University Press. 163 pp. Hardback $59.95, paper $21.95.

McMahon, April. 2000. Change, chance, and optimality. New York: Oxford University Press. 201 pp. Cloth $60.00, paper $19.95.

Merchant, Jason. 2001. The syntax of silence: Sluicing, islands, and the theory of ellipsis. New York: Oxford University Press. 262 pp. Hardcover $74.00.

Miller, Christopher Ray. 2000. La phonologie dynamique du mouvement en langue des signes québécoise. Saint–Laurent, Quebec: Champs Linguistiques, Fides. 176 pp.

Moro, Andrea. 2001. Dynamic antisymmetry. Cambridge, MA: The MIT Press. 142 pp. Paper $20.

Mukherji, Nirmalangshu; Bibudhendra Narayan Patnaid; & Rama Kant Agnihotri, eds. 2001. The architecture of language: Noam Chomsky. New York: Oxford University Press. 89 pp. $35.00.

Pintzuk, Susan; George Tsoulas; & Anthony Warner. 2001. Diachronic syntax: Models and mechanisms. New York: Oxford University Press. 380 pp. Paper $39.95.

Richards, Jack C. & Theodore S. Rodgers. 2001. Approaches and methods in language teaching. New York: Cambridge University Press. 270 pp. Hardback $59.95, paperback $21.95.

Warschauer, Mark & Richard Kern, eds. Network–based language teaching: Concepts and practice. New York: Cambridge University Press. 240 pp. Hardback $54.95, paper $22.95.

Wennerstrom, Ann. 2001. The music of everyday speech: Prosody and discourse analysis. New York: Oxford University Press. 317 pp. Paper $24.95. ■

NOTES ON LINGUISTICS

INTERNATIONAL LINGUISTICS COORDINATOR/EDITOR ■ *Michael Cahill*
ASSOCIATE EDITOR ■ *Eugene Loos*
CONSULTING EDITOR ■ *Alan S. Kaye*
FORMAT EDITOR / SECRETARY ■ *Betty Philpott*

INTERNATIONAL LINGUISTICS ADVISORS

Thomas Bearth ..... *U. of Zurich*
Steven Bird ..... *U of Pennsylvania*
Ruth Brend ..... *Michigan State U., r.*
Derek Fivaz ..... *Rhodes U., r.*
Peter Fries ..... *Central Michigan U.*
Shirley Brice Heath ..... *Stanford U.*
Yoshihiko Ikegami ..... *U. of Tokyo*
Frances Ingemann ..... *U. of Kansas*
Peter Ladefoged ..... *U.C.L.A.*
Sydney Lamb ..... *Rice U.*
W. P. Lehmann ..... *U. of Texas at Austin*
Howard P. McKaughan ..... *U. of Hawaii, r.*
Dinh Hoa Nguyen ..... *Southern Illinois U., r.*
John Oller ..... *U. of Southwestern Louisiana*
A. K. Pawley ..... *Australian National U.*
Richard Rhodes ..... *U. of California, Berkeley*
Malcolm Ross ..... *Australian National U.*
William J. Samarin ..... *U. of Toronto*
Hu Tan ..... *N.R.C. on Tibetology, Beijing*
Rudolph Troike ..... *U. of Arizona*
Richard Tucker ..... *Carnegie Mellen U.*
Robert Van Valin ..... *SUNY, Buffalo*
Ekkehard Wolff ..... *U. of Leipzig*

INTERNATIONAL LINGUISTICS CONSULTANTS

Rudy Barlaan ..... *Philippines*
John Bendor-Samuel ..... *U.K.*
Andy Black ..... *Mexico*
Cheri Black ..... *Mexico*
Regina Blass ..... *NEGST/Nairobi*
Les Bruce ..... *Dallas*
Donald A. Burquest ..... *U.T., Arlington*
Gene Casad ..... *Mexico*
Rod Casali ..... *Africa*
Ross Caughley ..... *Asia*
Lon Diehl ..... *Asia*
Warren Glover ..... *South Asia*
Kenneth J. Gregerson ..... *Asia*
Joseph Grimes ..... *Hawaii*
Austin Hale ..... *South Asia*
Lou Hohulin ..... *Dallas*
George Huttar ..... *Nairobi*
Connie Kutsch Lojenga ..... *Africa*
Stephen H. Levinsohn ..... *Colombia*
Ivan Lowe ..... *U.K.*
Steve Marlett ..... *Mexico*
Kenneth McElhanon ..... *Dallas*
Doris Payne ..... *U. of Oregon*
Tom Payne ..... *U. of Oregon*
Randy Radney ..... *Canada*
Gary Simons ..... *Dallas*
Keith Snider ..... *Canada*
David Weber ..... *Peru*
Klaus Wedekind ..... *East Africa*
Ursula Wiesemann ..... *Germany*

SIL International President ■ *Carolyn Miller*
SIL International Executive Director ■ *John Watters*
SIL International VP for Academic Affairs ■ *Karl Franklin*

NOTES FOR READERS ■ Permission is hereby granted by this journal to reproduce material from this issue, if readers obtain similar permission directly from its authors and if credit is given the authors and *Notes on Linguistics*.

NOTES FOR SUBSCRIBERS ■ *Notes on Linguistics* is a quarterly publication of the Linguistics Department of SIL International, designed to share information of practical, theoretical, administrative, or general interest. It is specifically intended to provide linguistic field workers with news, reviews, announcements, and articles that will stimulate interest in linguistics and help them stay current.

SUBSCRIPTION ■ Annual subscription rate: U.S address (delivered by U.S. Postal Service) $20.95; Outside U.S. (delivered by AIR) $23.95. A 20% discount is offered to resellers and personnel of SIL International. Prices include postage and handling. Subscriptions will be prorated so that all subscriptions end with the fourth issue of the volume.

INDIVIDUAL ISSUES ■ Individual copies of available issues are $3.00 each, plus a minimum of $2.75 handling (which generally covers about 3 issues). Not all back issues are still available.

ORDERING INFORMATION: Prepayment is requested from individuals; reseller and institutional purchase orders are accepted. Payment must be in U.S. funds, and checks must be drawn on a U.S. bank. Major credit cards are welcomed; to pay by credit card please include account number and expiration date. Electronic transfers are not accepted. Subscriptions, standing orders, and back issues are handled by: International Academic Bookstore, 7500 W. Camp Wisdom Road, Dallas, TX 75236, USA. Phone: 972-708-7404 FAX: 972-708-7363; Email: academic_books@sil.org

NOTES FOR CONTRIBUTORS ■ Readers are invited to submit publishable materials or letters of comment to the editor of *Notes on Linguistics* at The International Linguistics Center, 7500 W. Camp Wisdom Road, Dallas, TX 75236 USA. Internet e-mail: Mike_Cahill@sil.org (articles, etc.) or Eugene_Loos@sil.org (reviews). Intending contributors should ensure that their manuscripts conform to the standards laid down in the LSA Style Sheet printed periodically in the *LSA Bulletin*. Contributors who have copies of their manuscripts on magnetic media such as diskettes in PC/MS-DOS are urged to submit the same along with their manuscript. (3½" PC/MS-DOS diskettes or as e-mail attachments, in Word for Windows format are preferred.)

DISCLAIMER: Responsibility for the accuracy of quoted and referenced material rests solely upon the author of the item in which such material appears. Opinions expressed by the authors do not necessarily reflect policies of SIL.

ISSN 0736-0673

BEST COPY AVAILABLE

NOTES ON *LINGUISTICS*

VOLUME 4, NUMBER 4 2001

CONTENTS

COORDINATOR'S CORNER

A FAREWELL TO NOTES ON LINGUISTICS Michael Cahill..... 203

ARTICLE

GLOBAL LANGUAGE VIABILITY: CAUSES, SYMPTOMS AND CURES FOR ENDANGERED LANGUAGES Barbara F. Grimes..... 205

REVIEWS

THE PHONOLOGY OF GERMAN by Richard Wiese John M. Clifton..... 224

TOWARD A COGNITIVE SEMANTICS, I & II by Leonard Talmy Douglas Inglis..... 226

HOW THE BRAIN EVOLVED LANGUAGE by Donald Loritz Alan S. Kaye..... 230

COMPARATIVE SYNTAX OF BALKAN LANGUAGES by Maria-Luisa Rivero and Angela Ralli, eds. Julie May Kolgjini..... 233

EASTERN KAYAH LI: GRAMMAR, TEXTS, GLOSSARY by David Solnit Ken Manson..... 233

HANDBOOK OF NATURAL LANGUAGE PROCESSING by Robert Dale, Hermann Hoisl, and Harold Somers, eds. Mike Maxwell..... 236

CAMBRIDGE HISTORY OF THE ENGLISH LANGUAGE, IV, 1779–1997 by Suzanne Romaine, ed. Charles A. Mortensen..... 240

BANTU HISTORICAL LINGUISTICS: THEORETICAL AND EMPIRICAL PERSPECTIVES by Jean-Marie Hombert and Larry M. Hyman, eds. Steve Nicolle..... 247

THE ATHABASKAN LANGUAGES: PERSPECTIVES ON A NATIVE AMERICAN LANGUAGE FAMILY by Theodore B. Fernald and Paul R. Platero, eds. Gillian Story..... 254

MANDARIN CHINESE: AN INTRODUCTION by Mobo C.F. Gao Terrance R. Wardlaw, Jr..... 257

WHALES, CANDLELIGHT, AND STUFF LIKE THAT: GENERAL EXTENDERS IN ENGLISH DISCOURSE by Maryann Overstreet Richard L. Watson..... 260

SIL INTERNATIONAL
7500 W. CAMP WISDOM ROAD
DALLAS, TEXAS 75236 USA

NOTES ON LINGUISTICS

INTERNATIONAL LINGUISTICS COORDINATOR/EDITOR ■ *Michael Cahill*
ASSOCIATE EDITOR ■ *Eugene Loos*
CONSULTING EDITOR ■ *Alan S. Kaye*
FORMAT EDITOR / SECRETARY ■ *Betty Philpott*

INTERNATIONAL LINGUISTICS ADVISORS

Thomas Bearth ..... *U. of Zurich*
Steven Bird ..... *U.of Pennsylvania*
Ruth Brend ..... *Michigan State U., r.*
Derek Fivaz ..... *Rhodes U., r.*
Peter Fries ..... *Central Michigan U.*
Shirley Brice Heath ..... *Stanford U.*
Yoshihiko Ikegami ..... *U. of Tokyo*
Frances Ingemann ..... *U. of Kansas*
Peter Ladefoged ..... *U.C.L.A.*
Sydney Lamb ..... *Rice U.*
W. P. Lehmann ..... *U. of Texas at Austin*
Howard P. McKaughan ..... *U. of Hawaii, r.*
Dinh Hoa Nguyen ..... *Southern Illinois U., r.*
John Oller ..... *U. of Southwestern Louisiana*
A. K. Pawley ..... *Australian National U.*
Richard Rhodes ..... *U. of California, Berkeley*
Malcolm Ross ..... *Australian National U.*
William J. Samarin ..... *U. of Toronto*
Hu T'an ..... *N.R.C. on Tibetology, Beijing*
Rudolph Troike ..... *U. of Arizona*
Richard Tucker ..... *Carnegie Mellen U.*
Robert Van Valin ..... *SUNY, Buffalo*
Ekkehard Wolff ..... *U. of Leipzig*

INTERNATIONAL LINGUISTICS CONSULTANTS

Rudy Barlaan ..... *Philippines*
John Bendor-Samuel ..... *U.K.*
Andy Black ..... *Mexico*
Cheri Black ..... *Mexico*
Regina Blass ..... *NEGST/Nairobi*
Les Bruce ..... *Dallas*
Donald A. Burquest ..... *U.T., Arlington*
Gene Casad ..... *Mexico*
Rod Casali ..... *Africa*
Ross Caughley ..... *Asia*
Lon Diehl ..... *Asia*
Warren Glover ..... *South Asia*
Kenneth J. Gregerson ..... *Asia*
Joseph Grimes ..... *Hawaii*
Austin Hale ..... *South Asia*
Lou Hohulin ..... *Dallas*
George Huttar ..... *Nairobi*
Connie Kutsch Lojenga ..... *Africa*
Stephen H. Levinsohn ..... *Colombia*
Ivan Lowe ..... *U.K.*
Steve Marlett ..... *Mexico*
Kenneth McElhanon ..... *Dallas*
Doris Payne ..... *U. of Oregon*
Tom Payne ..... *U. of Oregon*
Randy Radney ..... *Canada*
Gary Simons ..... *Dallas*
Keith Snider ..... *Canada*
David Weber ..... *Peru*
Klaus Wedekind ..... *East Africa*
Ursula Wiesemann ..... *Germany*

SIL International President ■ *Carolyn Miller*
SIL International Executive Director ■ *John Watters*
SIL International VP for Academic Affairs ■ *Karl Franklin*

NOTES FOR READERS ■ Permission is hereby granted by this journal to reproduce material from this issue, if readers obtain similar permission directly from its authors and if credit is given the authors and *Notes on Linguistics*.

NOTES FOR SUBSCRIBERS ■ *Notes on Linguistics* is a quarterly publication of the Linguistics Department of SIL International, designed to share information of practical, theoretical, administrative, or general interest. It is specifically intended to provide linguistic field workers with news, reviews, announcements, and articles that will stimulate interest in linguistics and help them stay current.

SUBSCRIPTION ■ Annual subscription rate: U.S address (delivered by U.S. Postal Service) $20.95; Outside U.S. (delivered by AIR) $23.95. A 20% discount is offered to resellers and personnel of SIL International. Prices include postage and handling. Subscriptions will be prorated so that all subscriptions end with the fourth issue of the volume.

INDIVIDUAL ISSUES ■ Individual copies of available issues are $3.00 each, plus a minimum of $2.75 handling (which generally covers about 3 issues). Not all back issues are still available.

ORDERING INFORMATION: Prepayment is requested from individuals; reseller and institutional purchase orders are accepted. Payment must be in U.S. funds, and checks must be drawn on a U.S. bank. Major credit cards are welcomed; to pay by credit card please include account number and expiration date. Electronic transfers are not accepted. Subscriptions, standing orders, and back issues are handled by: International Academic Bookstore, 7500 W. Camp Wisdom Road, Dallas, TX 75236, USA. Phone: 972-708-7404 FAX: 972-708-7363; Email: academic_books@sil.org

NOTES FOR CONTRIBUTORS ■ Readers are invited to submit publishable materials or letters of comment to the editor of *Notes on Linguistics* at The International Linguistics Center, 7500 W. Camp Wisdom Road, Dallas, TX 75236 USA. Internet e-mail: Mike_Cahill@sil.org (articles, etc.) or Eugene_Loos@sil.org (reviews). Intending contributors should ensure that their manuscripts conform to the standards laid down in the LSA Style Sheet printed periodically in the *LSA Bulletin*. Contributors who have copies of their manuscripts on magnetic media such as diskettes in PC/MS-DOS are urged to submit the same along with their manuscript. (3½" PC/MS-DOS diskettes or as e-mail attachments, in Word for Windows format are preferred.)

DISCLAIMER: Responsibility for the accuracy of quoted and referenced material rests solely upon the author of the item in which such material appears. Opinions expressed by the authors do not necessarily reflect policies of SIL.

ISSN 0736-0673

BEST COPY AVAILABLE

From the Linguistic Coordinator

A farewell to *Notes on Linguistics*

This is the final issue of *Notes on Linguistics*. The reasons are partly practical, partly due to the changing world in which we live, but largely due to recognizing and dealing with the divided nature of our current audience.

The first issue of *Notes on Linguistics* (NOLx), in 1977, was explicitly labeled an 'in–house journal' for SIL members. As the years passed, more sophisticated articles also made their appearance, and scholars and libraries outside SIL expressed interest in our modest publication. The result was that NOLx eventually became a journal with two distinct, though overlapping, audiences. The first was the 'Ordinary Working Linguist', the SIL field worker who wanted practical tips on how to investigate languages. Articles aimed at this original audience had titles like 'What can we do with our data?' and 'A language lab in a soapbox'. As time went on, we started getting more descriptive articles such as 'The realis–irrealis distinction in Da'a' and more purely theoretical articles such as 'The relevance of rank and subordination to the application of phonological rules'. At this point, we were no longer aiming at a purely in–house audience, but also at the larger academic world. NOLx had become a journal with two largely divergent aims: a source of practical information to field workers, but also a source of language information and at least some engagement in theoretical linguistics.

Recently some of SIL's other Notes On journals have been stretched to the limit in terms of resources to produce their issues. This has caused us to reexamine our whole journal strategy, and we realized afresh that our two–audience journal discussed above was a bit schizophrenic. Fortunately, an alternative for all the Notes On journals was readily at hand. Since 1996, the Electronic Working Papers (http://www.sil.org/silewp/) has been a part of the SIL web site, and has featured a variety of original linguistic and other academic papers. While there have been a few papers posted every year, this site has thus far not come up to its potential. Our new strategy for publishing descriptions of languages and any other linguistics papers will focus on these Electronic Working Papers, encouraging SIL members and friends to submit their papers to this. Since EWP is a 'working papers' style of publication, authors will be free to further develop the papers and publish them elsewhere if they so choose. We expect a new website for publishing the types of book reviews which have been part of NOLx as well.

Thus in the internet–based academic world we live in, we judge electronic publication to be a better way to get our core academic production to a wider audience and at a cheaper cost as well. It is a step forward, and I am enthusiastic about our new publishing strategies.

For the kinds of papers that were purely the practical 'how-to' types of things, we anticipate a new in–house publication for SIL members that will deal with these matters. SIL members, you will be hearing about this soon! If a paper is judged to be of wider academic interest (and in these days of increasing interest in the investigation of endangered languages, methodological papers are becoming more attractive), we will encourage the author to submit it to Electronic Working Papers or another journal. But NOLx will be a thing of the past as of this issue.

I want to pay a special tribute to the men who preceded me as editors of NOLx, the International Linguistics Coordinators who also worked hard to produce NOLx. Karl Franklin was the founding editor in 1977, followed by George Huttar, Eugene Loos, and David Payne. Each of these talented individuals brought his own unique stamp to NOLx, and I am grateful to all of them for their contributions. As for me, I suppose it is less than a scintillating mark of distinction to be known as the editor who presided over the demise of the journal, but I trust that will not be my ultimate obituary.

For this final issue of NOLx, we are proud to present a major contribution by Barbara Grimes. For nearly 30 years, she has been the Editor of the Ethnologue, perhaps SIL's most enduring contribution to the linguistic world (http://www.ethnologue.com/). Her paper, 'Global Language Viability: Causes, Symptoms and Cures for Endangered Languages', brings the fruit of decades of research into one place. She provides an overview of many issues that have been increasingly discussed in the literature on endangered languages, and among whom SIL field workers often work. I hope you will enjoy it.

Michael Cahill
International Linguistics Coordinator
■

BEST COPY AVAILABLE

Global language viability:* Causes, symptoms and cures for endangered languages

BARBARA F. GRIMES
Ethnologue Editor 1971–2000
SIL International

1. Introduction. Of the 6,818 languages currently listed in the database of the *Ethnologue: Languages of the World* 330 languages each have one million speakers or more. This large size population contrasts sharply with the approximately 450 languages that are so small that they are in the last stages of becoming extinct, with only a few elderly speakers left in each one. At the same time, it may surprise many people to find out that the median size language in the world is 6,000 speakers; that is, half the languages in the world are spoken by 6,000 or more people, and half are spoken by 6,000 or fewer people.

2. Importance of Preserving Endangered Languages.

Prevent loss of culture and way of life. What does it matter if languages become extinct? Many people are alarmed when a plant or animal species becomes extinct. A **language dying** hits even closer to us; it **means that a unique creation of human beings is gone from the world**. Each language has grown up with its society, and is an expression of the facets of that society's culture. Each is an intricate system of words, phrases, clauses, and discourse patterns showing contrasts and agreements that its speakers use to describe their world and the customs they use in relating to each other. They use the language to tell their stories, recount their past, express their plans for the future, recite their poetry, and pass on their way of life.

Prevent loss of information about plants and wildlife. For example, when my husband, Joseph E. Grimes, was studying the plants and animals known to the Huichol people of Mexico, the Huichol were able to describe to him in their language many different plants and animals which they knew, including

---

*An English version of this paper was originally presented in Kyoto, Japan from Nov. 24–25, 2000, at a conference on Endangered Languages of the Pacific Rim. The conference was sponsored by The SCIENTIFIC RESEARCH OF PRIORITY AREAS under the Ministry of Education, Science, Sports and Culture of Japan under the direction of Professor Osahito Miyaoka formerly of the Graduate School of Letters, Kyoto University.

some animals for which the reference books did not give much information, like the jaguarundi (Grimes 1980). The Huichol told him what it looked like at different times of the year, what its habits were, and what it ate. They also told how they know if a carcass they find on the trail has been killed by a jaguar or a mountain lion. If he had not understood the Huichol language, there is no way they could have told him all that, because their Spanish was too limited to express it. These **treasure troves of information are lost each time a language dies**.

Preserve a people's identity. Many peoples feel that their **language expresses** their **identity**. A Russian linguist friend of ours, who was expelled from the USSR for political reasons in 1977, lamented to us that he could no longer write in Russian, his mother tongue, in which he had worked hard to be able to write well. His specialty was Russian lexicology, but he was more limited in expressing himself well in English or French, which were now his only two options.

3. Causes of language shift.

Parents push children to learn a prestige language thinking that they can only learn one language well. Linguists are aware that a language becoming extinct does not necessarily mean that the people who spoke it have all died. Instead, the speakers may shift to a different language, over one or more generations. Sometimes parents decide to not speak their mother tongue to their children, because they perceive an economic or educational advantage for their children in talking a second language. They do not want their children to suffer from not speaking that language well, as they did. And they mistakenly think that children can learn to speak only one language well. That is not true. Children can learn several languages well, as long as they know when to speak each one. Childhood is the best time to learn languages well.

Natural or man-made disaster—sudden shift. However, it does occasionally happen that a drought, famine, disease, war, flood, earthquake, tsunami, volcano eruption, acts of genocide, or other catastrophe wipes out a people. For example, the Paulohi language speakers in Maluku, Indonesia, experienced a severe earthquake and tsunami several years ago which killed all but about 50 of them (Ethnologue 502).

Migration outside of traditional territory—planned shift. Sociolinguists have found various other factors that seem to correlate with a language group shifting its language and diminishing in size. Voluntary or forced migration to a location outside their traditional territory has been a cause, as with Native American and Siberian peoples. The Naka'ela in Maluku, Indonesia

reportedly decreased in number after moving down from the mountains to a coastal town on Seram Island (Ethnologue:502).

Other language groups moving into a language's traditional territory has been a cause of loss, as with the Hawaiian people, who are now only about 20 percent of the total population in their traditional territory (Ethnologue:366).

The United States, Russian, and Canadian governments have moved indigenous children into boarding schools, insisting that they speak only the national languages in those schools. This resulted in those children being cut off from their traditional ethnic language and culture.

Use of second language in school—causing widespread shift. The language used in school for teaching can influence first or second language shift, as in the Philippines, where people educated before 1972 can speak English as a second language in addition to their home language, while people educated after 1972, when the language used for instruction changed to Tagalog, are more often fluent in Tagalog as a second language than in English. Some people now use Tagalog as their primary language who are not ethnically Tagalog.

National language policy—causing some but not universal shift. The desire to build a nation by a people and their government has contributed to language shift in the United States and Israel. When Tanzania became independent, the leaders decided to make Swahili their national language. For many years they limited linguistic investigation and promotion to Swahili. Many individuals became more bilingual in Swahili than they had been previously, and some now use Swahili as their primary language. But many people in the ethnic groups still speak their traditional languages as their primary language.

Factors working against language shift. People all over the world usually prefer their mother tongue over their second language because not only are they more fluent in it, but they speak it with the people they are closest to, and feel an affection for that language.

Other factors causing language shift. Urbanization, industrialization, other economic changes, government changes, and small population of the language being shifted from, have also been factors accompanying language shift. Because of the dynamic nature of these situations and other internal cultural factors, it has not been possible to predict language shift ahead of time for entire ethnic groups from any combination of these factors.

4. Effects of language shift (loss of mother tongue)

Dismay. The effect of these pressures and changes on the societies in question has been difficult and varied. My son, Charles Grimes, got together the last three speakers of the Kayeli language of Maluku, Indonesia, who were all over 60 years old, so he could take a word list in that language (Ethnologue:498; C. E. Grimes 1995, 2000; B. D. Grimes 1994). When he asked for certain words, none of the three speakers could remember them. Gradually a look of horror came over their faces as they realized that those words were lost forever. They pleaded with him to help them recover their language, but unfortunately he could only write down the words that they remembered.

Anti-social behavior and loss of self-esteem. Some people have said that they do not want to bring children into a world where their society, language, and people have no place. Some have turned to negative behavior like alcoholism, drugs, crime, or killing. The Waorani in Ecuador (Ethnologue:303), the Carabayo in Colombia (Ethnologue:293), and other groups in South America turned to killing, and for that reason some groups have still not had peaceful contact with the outside world. Or some have hidden from cultural pressures as have the Huichol in Mexico, who have moved farther back into the mountains. Or they have put up cultural barriers to make sure they do not accept any further innovations as the Cora in Mexico, who have appointed officials to keep innovations out. The Otomí of the Mezquital Valley in Mexico have acquired an inferiority complex about their culture and language through contact with the outside world (the Pan American Highway goes right through their territory). Many Otomí have turned to alcohol as an escape. Even something as simple as wearing dark glasses has been pointed to by anthropologists as a sign of looking for extra moral support by the Arapaho of the United States.

5. Symptoms of impending extinction of a language. For many of the 450 languages in the last stages of becoming extinct, the Ethnologue mainly has the information that they have a few elderly speakers left. But for others there is some additional information.

Drastic reduction of number of speakers, resulting in loss of aspects of culture. Students of languages and cultures in decline have noted that certain social, political, or religious functions of those societies must cease when there are not enough people left to handle them. Young people of marriageable age also have to look for mates outside their own group, and that presents the additional problem of what language to pass on to their children.

The Djingili in Australia have ten fluent elderly speakers of the language left, who use Kriol, an English–based creole of wider communication in Australia, as a second language (Ethnologue:717). The Gugubera in Australia have 15 fluent speakers left of their traditional language out of 50 members in the ethnic group, but the English of the ethnic group is poor (Ethnologue:718). The Guguyimidjir in Australia have twenty to thirty fluent speakers left out of 400 people in the ethnic group. There are 200 to 300 people in the ethnic group- who understand Guguyimidjir, but prefer Aboriginal English, including children (Ethnologue:719).

Changes in how the group makes a living. Hunter–gatherer groups ordinarily function well in small bands, but farming and pastoral societies do not. The Qawásqar are a small fishing band from southern Chile, 20 people, four of whom are monolingual. Reports are that the other Qawásqar are not very bilingual in Spanish. The youngest members are 3 to 20 years old. It is not clear how they can continue to function virtually monolingually unless they are able to maintain their livelihood and still keep to themselves, but they have made efforts to keep themselves physically isolated (Ethnologue:291)

Reduced domain in which language is used. Speakers of endangered languages may not have many people to use their own language with, so they use it with only family and close friends and possibly for ceremonies, as the Huilliche of Chile do (Ethnologue:291). The Hukumina language of Maluku, Indonesia had only one speaker left in 1989 who was eighty years old then. She could remember phrases and sentences, but had had no one to talk the language with for two decades (Ethnologue:497). Bardi children and adolescents in Australia seem to understand Bardi but never speak it (Ethnologue:715–716). The Eastern Ojibwa of Canada speak other varieties of Ojibwa to other Ojibwa people, and English with people who are not Ojibwa (Ethnologue:288). Some Australian Aborigines speak Kriol with other Aborigines, and English with other Australians. The Tlingit of Alaska and Canada all speak English as their first or second language. The youngest speakers of Tlingit are 60 to 65 years of age (Ethnologue:290).

Abandonment of mother tongue by young people. The Ili Turki language of China was just discovered by linguists in 1956. There are about 30 households of speakers, mainly older people. Younger people are intermarrying with people from other language groups; the group as a whole has gone over to Mandarin (Ethnologue:411).

BEST COPY AVAILABLE

The Manchu were once the rulers of China. Now there are 1,821,000 left in the ethnic group. But there are only an estimated 20 to 70 speakers left, all over 70 years old (Ethnologue:416).

The Mapia language in Irian Jaya, Indonesia is a Micronesian language linguistically, 180 miles north of the main coast of Irian Jaya. There is only one speaker left, who is elderly, and he still lives on Mapia Island. The rest of the ethnic group have moved to Micronesia, and speak another language (Ethnologue:478).

Attempts to maintain ethnic identity without using the mother tongue. The Itzá of Guatemala and Belize have only 12 speakers left, but they maintain their culture even though most of them now speak Spanish. However, it is not easy to fit a culture to a different language, and some features of the language are invariably lost when that is done (Ethnologue:307).

The Tariano people of the remote Vaupés River region of Colombia maintain their ethnic identity, but have no one left who speaks the language. However, across the border in Brazil, in the Vaupés River area of that country, in 1996 linguist Alexandra Aikhenvald found 100 elderly speakers of Tariano, an Arawakan language. There are 1,882 people in the ethnic group in both countries; those who do not speak Tariano speak Tucano, a Tucanoan language, or Nhengatú, a Tupí language—both languages of wider communication (Ethnologue:277, 298).

The Amahuaca people of Peru and Brazil speak a Panoan language. There are 300 in Peru and 220 in Brazil, but no children are learning the language. The group is disintegrating and losing its ethnic identity. Only 10 percent of the ethnic group can speak Spanish fairly well. Some speak only Spanish (Ethnologue:267, 347).

The Akurió people of Suriname are a hunter–gatherer people who speak a Carib language, and were just discovered by the outside world in 1969. For thirty years they had hidden in the tropical forest and their warm campfires were occasionally found by other indigenous people. All but one group now live with the Trió people, who speak a different Carib language. Akurio children now speak only the Trió language. The Trió respect the Akurió for their knowledge of the forest, but look down on them and their language. There were 44 or 45 Akurió in 1998 (Ethnologue:357).

The Arikara of the United States are reported to be one of the groups the well–known American explorers Lewis and Clark met in 1804 in North Dakota. At one time they had 30,000 people, but they were reduced to 6,000

because of smallpox. They now have only 90 speakers who are all middle–aged or older out of an ethnic population of 1,000. They have published instructional materials in the Arikara language, hoping that others in the ethnic group will learn the language (Ethnologue:362).

The Angosturas Tunebo of Colombia is a very small and endangered group. Some people think their speech is only a dialect of the Tunebo language because of the name, but it is actually a separate language related to the other Tunebo. Speakers of the four Tunebo languages have only limited understanding of each other's speech (Ethnologue:298).

Example of dialects of a language becoming extinct. A language variety becoming extinct is no less important just because it is a dialect of a larger language. There are cases where a dialect of a language is endangered, or has even become extinct, while other dialects remain viable. A dialect still needs linguistic description, and its speakers need literacy and other help.

Last speakers no longer know their own language very well. When languages are becoming extinct, the last few speakers sometimes do not speak the language well. They may simplify the grammar, change it to be more like their second language, or not know or use certain vocabulary.

Creole may develop if people from different languages come together. They also may speak their second language in a limited way. If they decide to speak their second language to their children when they speak it in a limited way, the children may end up not speaking any language like other native speakers. If an entire language group is like that, as on a plantation or in another situation where people from different language backgrounds are brought together suddenly and must communicate with each other, a new language like a pidgin, creole, or mixed language may develop.

Examples of endangered creole languages around the world. Among languages which are becoming extinct are different kinds of languages, like Chinook Wawa, a pidgin language in Canada and the United States. It is made up of words from Chinook, Nootka, Canadian French, and English (Ethnologue:282, 363). There are endangered creole languages, like San Miguel Creole French of Panama, which has Spanish influences (Ethnologue:344). Unserdeutsch is an endangered German based creole from Papua New Guinea (Ethnologue:793). Palenquero is an endangered Spanish based creole of Colombia whose speakers are mainly older, although 10 percent of those under 25 can speak it (Ethnologue:296–297). There are endangered mixed languages like Mednyj Aleut, a Russian–Aleut language from Russia which is assimilating rapidly into Russian

(Ethnologue:617), and Nguluwan, a Yapese–Ulithi language of Micronesia, which is assimilating rapidly into Yapese, according to linguist Osamu Sakiyama (Sakiyama 1980).

Examples of endangered sign languages. There are endangered deaf sign languages like Maritime Sign Language from Nova Scotia, New Brunswick, and Prince Edward Island of Canada (Ethnologue:286). Rennellese Sign Language from the Solomon Islands was invented by the first known deaf person from Rennell Island in 1915, and learned by all the 3,570 hearing people from there so they could talk to him (Ethnologue:804). The so–called 'Hawaii Pidgin' Sign Language came to light when a deaf person in Hawaii was involved in a traffic accident, and the court had to arrange a series of translators to interpret testimonies in court between that sign language, American Sign Language, and English. Providencia Sign Language of Providencia Island in the Caribbean is used by 19 deaf people and 2,500 to 3,000 hearing people. The deaf are fully integrated into that society. Since deafness is hereditary there, that sign language may actually not be endangered (Ethnologue:297).

6. Stages of language endangerment. At a colloquium (CLERD) in Germany in February 2000, linguists distinguished the following stages of language endangerment:

	STAGE	CHARACTERISTICS
1.	**Critically endangered.**	Very few speakers all 70 years old and older, great grandparent age.
2.	**Severely endangered.**	Speakers are only 40 years old and older, grandparent age.
3.	**Endangered.**	Speakers are only 20 years old and older, parent age.
4.	**Eroding.**	Speakers are some children and older people. Other children do not speak it.
5.	**Stable but threatened.**	All children and older people are speakers but few in number.
6.	**Safe.**	Not endangered. Language expected to be learned by all children and all others in the ethnic group.

Information needed on endangered languages: number of speakers compared to total number in ethnic group. It is important for us to have accurate information about each situation. Linguists have said that some of the most important information we need about each language is to know how many speakers are left, compared to how many people are in the ethnic group. Reasons for difficulty in getting accurate estimates need to be overcome.

I know how difficult it has been to get accurate estimates for the two languages where I have lived and worked for extended periods: Huichol of Mexico, and Hawaii Creole English of Hawaii.

Isolation due to lack of access to area. The Huichol live scattered out over a 50–mile by 50–mile area of very rugged mountains, in small family groups (Ethnologue:318). There have been no roads into the area until recently—only animal and foot trails, and a few airstrips.

Fear of outsiders leading to misinformation. The Huichol people do not give their correct names to strangers, because they believe that strangers will use those names to open them up to sorcery, causing them to have failing corn crops, get sick, or die. So they give false names to census takers and medical investigators. In addition, census takers and others do not necessarily get out to all the remote corners of the mountains where there is a house or two; the Huichol do not live in villages. When we began working there in 1952, estimates were of 7,000 or 8,000 Huichol. Later a school teacher who worked for the census bureau and was half Huichol himself estimated 12,000 speakers. I have since heard estimates of 20,000 and 51,000.

Language not considered worth spending money on. For the Hawaii Creole English language, Ian Hancock and Michael Forman, specialists in creole languages, have both estimated half the people in Hawaii to speak it, or 600,000, scattered out over the whole state (Ethnologue:366). But no one has ever done an accurate survey. At least twice the state legislature has requested one to be done, first of the Department of Health, and later of the Department of Education in 1989, but it was not done either time, probably because of lack of funding due to the low prestige of the language. People do not necessarily admit to speaking it because they are often laughed at, and accused of not speaking 'proper English'. So we stick with the estimates given by the creolists, which match our observations.

Spending enough time with people to observe their behavior. Other things linguists would like to know are if parents are passing on their mother tongue to their children, and if a second language is now used mainly for certain areas of life among its speakers, such as for home, school, government, religion, occupation, or commerce. If a linguist knows a language and has lived among its people, he or she is likely to have a much more accurate judgment about these things. The speakers are more likely to trust him and carry on their normal behavior when he is around. He can check what they tell him by observing what areas of life they use which language in, and what language they speak to their children. It is not easy to find out these things on a short stay, because the investigator cannot study causes, attitudes, cultural factors, or gain people's confidence in that short time.

Understanding the language used to describe the local language. Sueyoshi Toba of SIL reported to me that he recently was able to study all the reports that had been done on the Raute language of Nepal, including everything written in the Nepali language; but his report would have been much more incomplete if he had not understood the Nepali language.

Sometimes there is need to doublecheck census reports. One of the problems with the 1990 census of the USA was that Native Americans often reported that they spoke the language of their ancestors because they were of that ethnic group, even when they actually spoke only a few words or phrases of the language, or none of it. The same thing apparently happened in the 1990 census of Mexico, where an extinct language like Ópata, whose last speakers were previously reported to have died in 1930, was reported in the 1990 census to have a few speakers (Ethnologue:330). There are also remote languages like Yamana of Chile which is reported to possibly have more speakers out there somewhere (Ethnologue:292), and there are as yet uncontacted groups like the Pisabo in Peru (Ethnologue:351), and the Nukak Makú and Yarí of Colombia for which we have no speaker estimate (Ethnologue:296, 299).

It is important to have accurate information because if we claim that a language is not endangered when it is, those people and their language can miss the opportunity for linguistic and literacy help and description. If we claim that a language is endangered when it is not, then we lose our credibility, and cause people to lose interest in helping.

Need for in–depth studies of endangered languages. In cases where there are disagreements about language endangerment, it is important for there to be additional in–depth investigation by linguists, to find out what is accurate. I present the following information to show the need for more in–depth

investigation, and also to illustrate some of the reasons why it is difficult to get accurate reports.

There are occasional cases of misinformation that need to be balanced by more accurate reports. One preliminary endangered languages report says that the Lacandón in southern Mexico, a Mayan language, have no speakers younger than 30 years old. However, Mary Baer of SIL, who has lived among them since 1942 and speaks their language, says there are currently nearly 1,000 speakers of all ages who talk to her in Lacandón. That is the entire ethnic group.

Decline may not mean death if proper measures are taken in time. The preliminary report says there are no Mayo language speakers younger than 40 years old. This is a Uto–Aztecan group of northern Mexico. But Larry Hagberg of SIL, who has lived among them for years and speaks their language, says there are over 100 Mayo villages, and native speakers report that half a dozen of the villages have speakers under 40 years old. He has met many fluent speakers under 40, and agrees with the 1995 census figure giving 40,000 Mayo who still use the language in the home, including a few children. Hagberg says the Mayo people are very shy, particularly concerning their own identity as Mayo. The more monolingual people tend to avoid contact with outsiders. Those who are bilingual often deny any knowledge of the Mayo language when asked by an outsider if they speak Mayo. He thinks the language may be used more vigorously now than when he first arrived there because the government is doing things to stimulate pride in their ethnic heritage, like giving generous scholarships to Mayo young people who can demonstrate a certain degree of proficiency in the language. There is also a radio station that broadcasts in Mayo.

Bilingualism may not lead to extinction of the mother tongue. A preliminary endangered language report for Guatemala says the Lanquin Q'eqchi' children are abandoning their language. However, Francis Eachus of SIL, who has lived there since 1955 and speaks the language, says all the people and their children there speak the language, even though they are bilingual in Spanish. Bilingualism does not necessarily lead to language loss. There are other interesting cases where reportedly endangered languages are surviving.

Guatemala. Uspanteko is on the preliminary list for Guatemala as having the children abandoning it, but Stan McMillen of SIL, who has lived there for years and speaks the language, says there are 3,000 speakers, and children in their main center of Las Pacayas speak the language. The

language will be broadcast over the radio soon, and the Mayan Academy is actively promoting the use of the language.

Sipakapense is on the preliminary list for Guatemala as small, but Ed Beach and Eric Kindberg of SIL recently found 8,000 speakers who maintain the use of Sipakapense, out of over 12,000 in the ethnic group, and many children actively using the language. There is a strong sense of identity with and ownership of Sipakapense. Most people did not learn to speak Spanish until they began attending school, and they still understand Sipakapense better than Spanish.

Tekiteko is on the preliminary list for Guatemala as small, but it has 1,000 to 1,500 speakers, according to Ed Beach, who has lived there since 1978 and speaks the language. He says the larger estimates of 5,000 given 20 years ago were not accurate, but the size was 1,000 to 1,500 at that time and that size has been maintained. There is a resurgence in use of the language. In the largest villages Tekiteko is used commonly by children, but in some other villages parents push their children to speak only Spanish.

Cunen K'iche' is on the preliminary list for Guatemala as small, but it has 9,000 speakers, according to Larry Marhenke who has lived there since 1991. This includes 2,000 who live in Guatemala City but maintain ties in the villages. Over 90 percent of the people regularly speak their language. A few children are not being taught the language but most children can be heard playing in it. It is the dominant language heard in the local market. There is also a significant degree of monolingualism among men and women in the countryside.

Peru. A preliminary report on endangered languages of Peru lists Cajamarca Quechua with 10,000 speakers. David Coombs of SIL, who has lived there for 19 years and speaks the language, says that he tried to get an accurate estimate when doing his dissertation research in the late 1970s, but the very low prestige of the language and the fairly high level of bilingualism in Spanish made that impossible. For example, a few young male bilinguals told him and government census takers that only a few very old people still spoke Quechua, even though their own female family members later proved to be monolingual Quechua speakers, including young women. Coombs estimates that 30,000 to 35,000 people living in at least 28 villages and hamlets speak Quechua. But he says the pressure on Quechua is severe, and the language is relatively strong in only 8 to 10 communities. Very young children in some areas are still learning it.

The preliminary endangered languages report on Peru lists 350 to 450 Sharanahua. Eugene Scott of SIL has lived there since 1958, and speaks the

language. He estimates the population is closer to 450 to 500, plus others in Brazil who go by other names. Those in Peru speak Sharanahua exclusively in their homes.

The preliminary endangered languages report on Peru lists 890 to 2,500 Matsés. Harriet Kneeland and Harriet Fields of SIL have lived there since 1969, speak the language, and say there are 2,500 to 3,000 in Peru and Brazil, nearly all of whom speak Matsés. On the Peru side the school teachers are bilingual Matsés speakers. Some Matsés speak Spanish or Portuguese as second language.

The Philippines. A preliminary report on endangered languages of the Philippines lists the Central Cagayan Agta language as having 500 speakers, who represent 40 to 60 percent of the ethnic group, and a few children. The second language is used for all important functions except perhaps family. The people are mildly supportive toward their language, and the language is endangered according to the report. However, Roy Mayfield of SIL, who has lived there since 1959 and speaks the language, says there are 700 to 800 speakers, who represent 95 percent of the ethnic group and all or most of the children. Speakers use Agta with all Agta people and even with some Ilocano, Itawit, and Ibanag people who have learned Agta. They use Ilocano with all other people who are not Agta. Though government has decreed that Tagalog is to be used in education, the Agta are strongly supportive of their own language and there is a strong identity of the Agta people with the Agta language.

Indonesia: Irian Jaya. A preliminary report on the endangered languages of Irian Jaya, Indonesia, lists Auye as having 300 speakers who represent 40 to 60 percent of the ethnic group, and a few children still speaking the language. It reports that the second language is being used in all areas of life except perhaps the family circle, that the speakers have a neutral attitude toward their language, and it is potentially endangered. However, Mike Moxness of SIL, who has lived there and speaks the language, says Auye should not be on the endangered list. Linguists are still contacting part of the Auye group. They have contacted 350 of them so far, and he doesn't know any person, including children, who does not speak the language. Most of the speakers do not know Indonesian.

The Foya or Abinomn language of Irian Jaya was also on the preliminary endangered languages list as having possibly 20 speakers, which represents 30 percent or less of the ethnic group, and no children that speak the language. It was reported that the second language is used for all areas of life except perhaps the family, and that people have negative or neutral

BEST COPY AVAILABLE

attitudes toward their language, and it is potentially endangered. However, Mark Donohue of the University of Sydney has done fieldwork there and he says there are 300 speakers. They have a low level of schooling but are enthusiastic about literacy, want their own dictionary, and want to write stories in their language.

The Tobati or Yotafa language of Irian Jaya was on one preliminary endangered language list as having 2,460 speakers and being endangered. But it was on another preliminary endangered language list as having 100 speakers, and being seriously endangered. Donohue says there are 350 speakers.

The Yoke or Yoki language of Irian Jaya was on one preliminary endangered language list as having possibly 20 speakers, who represent 20 to 60 percent of the ethnic group and none or a few children as speakers. It was reported that the second language is used for all areas of life except perhaps the family, and that the people have a negative or neutral attitude toward their language, and that it is seriously endangered. However, Donohue says there are 200 speakers of all ages, language use is vigorous, and the people know hardly any Indonesian. They know some Warembori, a nearby language, but the Warembori people know even more of the Yoke language.

Kowiai of Irian Jaya was on one preliminary endangered language list with 500 speakers, who represent 40 to 100 percent of the ethnic group, and a few to half of the children. It was reported that the second language was used for all or most areas of life, people had a neutral attitude toward their language, and it was potentially endangered. However, Roland Walker of SIL who lived among them for several years says there are 600 people, nearly all of whom speak the language, including children. The second language is being used for few or no areas of life, and the people strongly support their language.

Sulawesi. A preliminary endangered languages report on Sulawesi, Indonesia, lists Panasuan as having 800 speakers, who represent 40 to 60 percent of the ethnic group, and 30 to 50 percent of the children. Panasuan is reported to be used for some important areas of life, the people are reported to be neutral toward using their language, and it is potentially endangered. However, Sem, a man from Kalumpang, recently visited there, where he has friends, and he says there are 900 or more speakers, all the ethnic group including children speak the language. The second language isn't dominant in any areas of life, and the people are strongly supportive toward their language.

The Talondo' language of Sulawesi was also on the preliminary list, with 400 speakers, who represent 40 to 60 percent of the ethnic group, and a few to half of the children. The second language was reported to be used for all areas of life except perhaps the family, people were reported to be neutral toward their language, and it was potentially endangered. However, Sem, the man from Kalumpang, has also recently visited the Talondo' people where he also has friends. He says there are 500 speakers, all the ethnic group speaks the language, including most children. Very small children know only Talondo'. They use Talondo' in the home, for sports, and children playing. Indonesian is used for trading because the sellers are Bugis people. They use Kalumpang with the Kalumpang people. School teachers are not Talondo', so Talondo' is not used in school.

The Napu language of Sulawesi was also on the preliminary list, with 4,000 speakers, who represent 40 to 60 percent of the ethnic group, 30 to 50 percent of the children. The second language was reported to be used for most of the important areas of life, people were reported to be neutral toward their language, and it was potentially endangered. However, Roger Hanna of SIL has lived there for over 10 years, and says the 2000 census information gives 6,000 speakers, the percentage of children who speak the language is over 60 percent, Napu is used for some key roles in life. He says Napu people are at least mildly supportive of their language. Interest in Napu literature is high.

Papua New Guinea. For Papua New Guinea, one preliminary report on endangered languages listed Mussau-Emira with 3,650 speakers, and says it is potentially endangered. Another report says there are 3,500 speakers, 40 to 60 percent of the ethnic group speak the language, including a few children. It reports that the second language is used for most of the important areas of life, people are negative or neutral toward their language, and it is endangered. However, John Brownie of SIL, who lives there, says he estimates the speakers as 3,500 to 5,000; a new census was taken last July but the report is not out yet. He says of the 3,500 resident population, virtually everyone is a speaker of the language including children, but for a large group outside the traditional area in Kavieng town, it is probably not universal. He said there was a period when people were using more Tok Pisin, their second language, with children, but that is past, and people are now passing on their language to their children. He says the language is very healthy, and it is being used for all the key roles in the villages. He says the success of elementary schools in the local languages in Papua New Guinea in the last five years has reinforced positive attitudes and he does not see anyone with negative attitudes.

One preliminary endangered language list for Papua New Guinea lists Kakabai with 880 speakers. However, Russ Cooper of SIL says population estimates varying from 700 to 1100 are given but the situation is confusing because the people move their entire villages, sometimes to a different census area or government council jurisdiction. Anyone school age or older seems to have some fluency in at least three languages, and the average is five languages. But Kakabai is the language that keeps the group's culture and identity together.

On the other hand, Gweda or Garuwahi is a language of Papua New Guinea on one preliminary list given with 225 speakers, and on another preliminary report with 200 speakers, but Russ Cooper says it was found on a recent survey to have only members of one family, fewer than 30 to 40 people, including an older couple whose children and grandchildren are the only ones who speak Gweda with them. The rest of the ethnic group speak Taupota or Tawala.

Wataluma was on one preliminary Papua New Guinea list as having 190 speakers who represent 70 to 100 percent of the ethnic group, and 60 to 100 percent of the children. Wataluma was used for some or most important functions, the people were mildly supportive toward their language, and it was potentially endangered. Russ Cooper says he talked at length with a village headmaster from Wataluma who feels it is a viable language. Cooper has heard other Papua New Guineans also make the claim that Wataluma is alive and well, though a very small group.

Magori was on a preliminary Papua New Guinea list as endangered with 200 speakers who represent 40 to 60 percent of the ethnic group, and a few to half the children. The second language was reported to be used for all important functions, the people were negative to neutral about Magori. Russ Cooper has heard others from Papua New Guinea make the claim that Magori is not endangered. He says, if Magori is surviving, and holding its own, that would be worthy of study because 'we've been reporting its near demise for decades now'. He agrees that one thing that has happened to turn language attitudes around in Papua New Guinea has been the government thrust toward elementary education in the local languages.

8. Cure for endangered languages. There are some groups in the world who, with help from linguists, are attempting to increase their speaker population by publishing pedagogical grammars with audio tapes, dictionaries, newspapers, having courses taught in schools and community colleges, evening classes for adults, and programs on the radio and television. These are having some success. But the most important thing

that can happen to preserve a language is for parents to continue passing on their mother tongue to their children in the home.

Hebrew. There is some good news. There are languages that have been able to recover wholly or partially from being endangered. The most famous example of all, of course, is Hebrew, which is now estimated to have 5,150,000 mother tongue speakers in the world. They make up 81 percent of the population of Israel. Nearly everyone in Israel speaks Hebrew as either their first or second language (Ethnologue:536).

Cornish (Great Britain). Cornish from the United Kingdom is another example of a language which is coming back. It became extinct as a mother tongue in 1777, but members of the ethnic group have brought it back recently so that there are reported to be a number of people under 20 who use it as first language, 1,000 who use it as their everyday language, and 2,000 others who speak it fluently. They have evening classes, correspondence courses, summer camps, childrens' play groups, and it is taught in some schools (Ethnologue:707).

Hawaii. Immersion schools and Hawaiian studies program at the University of Hawaii present some interesting facts. In 1778 when people from outside Polynesia first came to Hawaii there were estimated to be over 500,000 Hawaiians, all of whom spoke Hawaiian. In 1900 there were an estimated 37,000 mother tongue speakers of Hawaiian but today there are only 1,000 mother tongue speakers left. Half of the mother tongue speakers are elderly. There are 8,000 others who can speak and understand it out of 220,000 members of the ethnic group. In the early days many of the Hawaiians died from measles and other diseases brought in by foreigners, for which they did not have immunity. When the Hawaiian monarchy was overthrown by American businessmen in 1893, English was made the national language, and people were told to not talk Hawaiian at home or in school. This policy continued when the United States annexed Hawaii in 1898. However, recently the Hawaiians have started total immersion schools, adding one school year each year, and now they have grades one through twelve taught in the Hawaiian language. The total immersion schools are growing in number. Many parents are also studying the language so they can talk to their children at home in Hawaiian. The University of Hawaii now has a graduate program in Hawaiian Studies, which includes a concentration in the Hawaiian language (Ethnologue:366–367).

New Zealand. The Hawaiian immersion schools were patterned after those of the Maori in New Zealand, a related Polynesian language, which are having similar results. There are now an estimated 50,000 to 70,000 mother

tongue speakers of Maori, and 100,000 other Maori people who understand but do not speak it, out of 310,000 people in the ethnic group (Ethnologue:739).

Mexico. There are other languages that have been able to come back through literacy classes and other education, linguistic work, and modern medicine. Two examples are from Mexico. In 1942 there were an estimated 178 Lacandón, a Mayan language in southern Mexico already mentioned, but in the year 2000 nearly 1,000 speakers (P. Baer 1998; M. Baer 2000). In 1951, there were 215 Seri, a Hokan language in northern Mexico, but in 2000 it had an estimated 800 speakers (Moser 1998, 2000).

Ecuador. Another example is from Ecuador. The Waorani were estimated to be 150 in number when they became widely known for killing five missionaries in 1956. They were declining at that time because of repeatedly feuding and killing each other, including their own family members. Now after linguistic work, literacy classes and other education, modern medicine, and a consensus among them not to seek revenge, they are estimated to be nearly 1,400 in number (Smith 1998).

Brazil. The Nambikuara of Brazil were reduced in the 1940s from 10,000 to 600 through measles but have been experiencing renewal through literacy and other education, medicine, and linguistic work. They are now estimated to be 900 in number (Kroeker 1998).

8.1 Other endangered languages via research of outside linguists, anthropologists and physical science people. Linguists who have published grammars, dictionaries, and annotated text collections have described some endangered languages. Some have also had their cultures and ecology described by anthropologists, botanists, zoologists, and other scholars, but many endangered languages are still needing these kinds of studies.

9. Conclusion. There is hope for endangered languages if increased effort is put into in–depth language evaluation, linguistic description, literacy work and other education, access to modern medicine, sanitation, agriculture, and other forms of aid.

References

Baer, Mary. 2000. Personal communication.

Baer, Sr., Phillip. 1998. Not the last of the Lacandones. In Back from the Brink: Sixteen Accounts of Ethnic Renaissance by Richard S. Pittman. Waxhaw, North Carolina: SIL International, pp. 15–16.

CLERD. 2000. Colloquium on Language Endangerment, Research and Documentation: Setting Priorities for the 21st Century, held at Bad Godesberg, February 12–17, led by Matthias Brenzinger, Köln University.

Ethnologue. 2000. Barbara F. Grimes, ed. Ethnologue: Languages of the World. Dallas: SIL International.

Grimes, Barbara F. 2000. Ethnologue: Languages of the World. Dallas: SIL International.

Grimes, Barbara D. 1994. Cloves and nutmeg, traders and wars: language contact in the Spice Islands. In Language contact and change in the Austronesian world, Trends in Linguistics: Studies and Monographs 77, Thomas Dutton and Darrell T. Tryon, eds. Berlin: Mouton de Gruyter, pp. 251–274.

Grimes, Charles E. 1995. Digging for the roots of language death in eastern Indonesia: The cases of Kayeli and Hukumina, published as ERIC Clearinghouse on Languages and Linguistics, Document No. ED 384231, Washington: U.S. Department of Education, paper presented at Linguistic Society of America meetings, New Orleans.

____. 2000. Defining speech communities on Buru Island: A look at both linguistic and non–linguistic factors. In Spices from the East: Papers in languages of eastern Indonesia, Charles E. Grimes, ed. Pacific Linguistics 503:73–103.

Grimes, Joseph E. 1980. Huichol life form classification I: Animals. Anthropological Linguistics 22:187–200; II: Plants. Anthropological Linguistics 22:264–74.

Kroeker, Menno. and Barbara Kroeker. 1998. No longer endangered Nambikuara. In R. S. Pittman, pp. 33–34.

Moser, Mary Beck and Richard Pittman. 1998. Sonoran desert survivors. In R. S. Pittman, pp. 11–12.

Moser, Mary Beck. 2000. Personal communication.

Pittman, Richard S. 1998. Back from the brink: Sixteen accounts of ethnic renaissance. Waxhaw, North Carolina: SIL International.

Sakiyama, Osamu. 1980. The characteristics of Nguluwan from the viewpoint of language contact. In Islanders and their outside world, Machiko Aoyagi, ed. Tokyo: Rikkyo University.

Smith, Donald E. 1998. Waorani renaissance. In Pittman, R. S. pp. 6–7.

Barbara F. Grimes, 84664 Ala Mahiku Dr. #191B, Waianae, HI 96792 USA.
E-mail: Barbara_Grimes@sil.org.

■

BEST COPY AVAILABLE

Reviews

The phonology of German. By RICHARD WIESE. 1996. The Phonology of the World's Languages. Oxford: Clarendon Press. 351 pp. Hardback $71.26, paper $27.65.

Reviewed by JOHN M. CLIFTON
SIL—North Eurasia Group and University of North Dakota

'Don't judge a book by its title'.

If that isn't an adage, maybe it should be. My first reaction to this book was negative, but after some contemplation I concluded that was due to expectations based on the title. Which in turn leads me to some observations on the organization of and the intended audience for a phonological description.

There are at least two major ways to organize the phonological description of a particular language; there are three major audiences for which a phonological description may be intended. And the relationship between the organization of a phonological description and the intended audience is not arbitrary.

One way to organize a phonological description is around the structure of the language itself. That is, one begins by analyzing the phonological alternations associated with some small part of the language, and then incrementally adds additional forms and their analyses. A second way is to organize it around theoretical models. When a description is organized around theoretical models, a given paradigm or alternation might well be treated in various sections of the presentation.

Considering possible audiences for a phonological description, one audience consists of theoreticians who want to see how particular models can account for the facts of a given language. A second audience is scholars of the language who are interested in what insights the models give into the facts of the language. A third audience includes those who are not experts in the language, but who want to learn about how the language works within the rigor provided by explicit models. The primary goal of this third audience is to come away with an overall understanding of the 'genius' behind the phonological system of the language. And most SIL members are writing for this third audience when presenting the system of a little–studied language.

BEST COPY AVAILABLE

For the first audience, a description organized around theoretical models is undoubtedly best. For the second audience, either type of description will probably work, although the writer will need to spend extra time presenting the model clearly. My experience is that for the third audience, a description organized around theoretical models will probably result in confusion, not enlightenment. The reader is unlikely to come away from such a description with a clear, overall understanding of the phonological system.

In this light, I can finally explain my initial reaction. From the title, I expected the book to be first and foremost a description of the phonological system of German. It is not. Instead, the book is a good example of a description organized around a set of interrelated theoretical models.

Wiese is clear right from the beginning that the major audiences are the first and second groups outlined above: 'the linguist (especially the phonologist)', and 'teachers and students of German, especially at the college and university level' (pp. 2–3). The book has two goals: to give 'a thorough and in–depth survey of the phonological system of present–day German' and to show the 'application of recent theories and models as developed in phonology to the German language' (p. 1). I feel the book meets its second goal better than it does its first; the organization around theoretical models tends to obscure the overall system of the language itself. Similarly, I feel that teachers and students of German will have a difficult time with the book unless they have a reasonably good familiarity with generative phonology.

The book is arranged into nine chapters. After a short introductory chapter, the second chapter presents the phonemic system. In addition to proposing a set of distinctive features, W presents a number of problems that arise in standard phonemic analysis and promises to deal with them in later chapters.

Chapters 3 through 5 are highly interrelated. Prosodic structure is the focus of chapter 3. Well over a third of the chapter deals with the syllable; in addition W discusses feature geometry, the foot, the phonological word, the phonological phrase, and the intonational phrase. Following on from this, the focus of chapter 4 is the application of the theory of Prosodic Morphology—that is, instances in which 'the wellformedness of complex words crucially depends on prosodic conditions being met' (p. 85). Finally, in chapter 5 W deals with the application of the theory of Lexical Phonology to German.

In chapter 6 W moves back to feature representation and proposes an analysis of phonological features in terms of underspecification theory. Then, in chapter 7 W systematically presents analyses of various morphophonemic alternations and allophonic variation within the general

framework outlined in chapters 3 through 6. Included in this chapter is a discussion of the sonority hierarchy. Word, compound, and phrase stress is the focus of chapter 8. The book closes with a two–page chapter of concluding remarks and includes areas for further research. According to Parrott (2001), the paperback edition of the book (Wiese 2000) also includes a five–page postscript that deals with Optimality Theory.

If you are looking for a book to give you a handle on how German works, this book is probably not for you. And I don't feel it presents a good model for SIL field workers who are interested in organizing the phonological description of some little–studied language. But as indicated above, it does apply a large number of interrelated phonological models to the facts of the German phonological system. This book should be of interest to those who want to see how various models fare in explaining a large body of data. It represents a good attempt at showing how the models complement each other to deal with a broad range of phenomena.

References

Parrott, Jeffrey K. 2001. Review of The phonology of German. Linguist List 12.2354. http://www.linguistlist.org/

Weise, Richard. 2000. The phonology of German. Oxford: Oxford University Press.

John M. Clifton, 511 Ave. of Americas, Suite 15, New York, NY 10011;
E-mail: debbie–john@sil.org

Toward a cognitive semantics, I and II. By LEONARD TALMY. Massachusetts Institute of Technology. 2000. Vol. 1, 565 pp.; Vol. 2, 495 pp. Both hardcover volumes $110.

Reviewed by DOUGLAS INGLIS
SIL—Mainland Southeast Asia Group

This two–volume set represents a journey of exploration within cognitive semantics in the direction of an account of conceptualization that seeks to unify general psychological structures (thus 'cognitive') with linguistic structures. This is a mass collection of Talmy's thinking articulated throughout his writing career. These volumes extend beyond mere collection as they also offer updates, revisions, and embellishments to the original publications spanning about twenty years. The papers have also been arranged in a topical rather than a chronological manner. The cognitive linguistic

paradigm has been summarized nicely by Dirk Geeraerts, the current president of the International Cognitive Linguistics Association (1997:8):

> Cognitive Linguistics is the study of language in its cognitive function, where *cognitive* refers to the crucial role of intermediate informational structures in our encounters with the world ... [Cognitive linguistics] focuses on natural language as a means for organizing, processing, and conveying that information. Language, then, is seen as a repository of world knowledge, a structured collection of meaningful categories that help us deal with new experiences and store information about old ones.

In the spirit of Geeraerts, these two volumes contribute foundationally to the discussion of the nature of linguistic representation of conceptual structure as outlined in cognitive linguistics. Volume I explores the foundational systems by which language shapes concepts, while Volume II investigates fundamental typologies and processes that structure concepts.

The two volumes represent a progression of conceptual ideas that start in language structure and proceed to nonlinguistic cognitive systems. Volume I deals largely with 'schematization' and outlines several systems that structure concepts, namely, 'configurational structure', 'distribution of attention', 'force and causation'. It is divided into four parts. Part 1 concerns the relationship of grammar to cognition and sets the theoretical tone for the two volumes. Part 2 discusses the configurational structures of fictive motion and space. Language instances such as, 'This fence goes from the plateau to the valley; the cliff wall faces toward the island', in which motion is projected onto an immovable object are compared with visual instances such as, successive flashes along a row of light bulbs, as on a marquee, and the perception of a plus sign as involving the sequence of a vertical stroke followed by a horizontal stroke. The framework developed is from an OVERLAPPING SYSTEMS model of cognitive organization, which briefly states that each cognitive system has some structural properties that may be uniquely its own, some properties that it shares with only one or a few other cognitive systems, and some that it has in common with all the cognitive systems. This view bears more cognitive integration with other cognitive systems than pictured by a strict modularity notion (e.g. Fodor 1983), and in my opinion suggests that linguistic frameworks which take this integration as a working assumption are better equipped to handle more complex issues in language such as metaphor, idiom, fictive motion, etc. Part 3 concerns an in–depth look at cognitive **ATTENTION** which figures foundationally in many cognitive linguistic descriptions today (Langacker 1991; Tuggy 1993; van Hoek 1997; Inglis to appear). For example, figure and ground orientation as a sub–topic of cognitive attention receives a psycholinguistic discussion in Chapter 5. Figure and ground are

fundamental notions in linguistics used to describe such diverse phenomena as lexical reference, grammatical head constructions, and transititvity (see Langacker 1987). Finally, Volume I concludes with Talmy's watermark in linguistics, FORCE DYNAMICS. This discussion entails force as a fundamental notion in the structure of language and motivates causation in this structure. Force dynamics comprises a framework of causation that encompasses also 'letting', 'hindering', and 'helping', among other concepts not usually included in the causation context. It further helps to motivate the grammatical category of modals along with grammatical elements.

Volume II is composed of three parts. Part 1 presents lexicalization patterns and event structures. Lexicalization studies the systematic relations in language between meaning and surface expression (overt linguistic form). By isolating elements within the separate domains of meaning (e.g. motion, path, figure, ground, manner, cause) and overt expression (e.g. verb, adposition, subordinate clause, etc.), then by mapping relationships between meaning and form, a range of universal principles and patterns arise. The concern is how the patterns of mapping compare cross–linguistically. This procedure then culminates in the consideration of cognitive processes and structures that might give rise to these typologies and patterns. The chapters in part 2 move beyond typology in focusing on online interactive processing of multiple factors. This section analyzes, for example, regular linguistic situations in which a portion of discourse received by an addressee provides two or more specifications for the same referent. When these situations are in conflict certain cognitive operations resolve the conflict. The nature of these cognitive systems is the target of the discussion. The topics of this section include semantic shifts, blends, juxtapositions, juggling, and blocking. The interaction then broadens to analyze the cognitive interaction within the goals of communication. This considers communication from the perspective of the psychological situation during the course of communication. In part 3, extensions of the conceptual and cognitive analyses applied to language in preceding chapters are oriented to more non–linguistic cognitive systems that underlie culture and narrative.

Talmy's collection represents sound theoretical articulation that is all encompassing and far reaching for field linguists. Although theoretical in orientation, at each stage of the composition a lucid linguistic analysis provides the descriptive direction that a cognitive theory implicates. This direction should interest linguists that view language holistically in its relation to anthropology, culture, and psychology. These two volumes are as much a plea to consider cognitive linguistics as a valid enterprise (in contrast to a generative paradigm) for grammatical descriptions and lexicography work. For example, in my own descriptive work in Southeast Asian

BEST COPY AVAILABLE

languages, following Langacker (1991), I have posited a semantic structure for numeral classifiers that allows one to better describe the grammatical phenomena surrounding these constructions. I claim numeral classifiers instantiate the noun into an experiential domain necessary for reference in communication. This analysis accounts for aspects of the numeral classifier that up to now remain ambiguous, e.g. polysemy between a noun and classifier (Inglis to appear). Schematization processes, such as that in Talmy's Volume I Chapter 3, contribute theoretical foundation to my description. Also, in my work in lexicography, intuitive decisions made in sense discriminations for some nominalization and verbalization processes, such as in Newell (1995:200–205), find a solid theoretical framework in cognitive linguistics (Langacker 1987; Inglis 2000); Talmy's Volume I, Chapter 5 'Figure and ground in language' discusses constructs central to these intuitive sense discriminations. Talmy highlights the path taken by cognitive linguistics and offers to us perspectives and assumptions on language that underlie frameworks worthy of our descriptive energies.

REFERENCES

Fodor, Jerry. 1983. Modularity of mind: An essay on faculty psychology. Cambridge, MA: MIT Press.

Geeraerts, Dirk. 1997. Diachronic prototype semantics: A contribution to historical lexicology. Oxford: Clarendon Press.

Inglis, Douglas. 2000. AL 642 Semantics. Graduate School, Linguistics. Payap University, Chiangmai, Thailand. Unpublished course curriculum.

_____. To appear. Conceptual structure of numeral classifiers in Thai. In Cognitive Linguistics in non–Indo European languages, Eugene Casad and Gary Palmer, eds. Berlin, Germany: Mouton de Gruyter.

Langacker, Ronald W. 1987. Foundations of cognitive grammar, 1: Theoretical Prerequisites. Stanford: Stanford University Press.

_____. 1991. Foundations of cognitive grammar, 2: Descriptive Applications. Stanford: Stanford University Press.

Newell, Leonard E. 1995. Handbook on lexicography: for Philippine and other languages. Linguistic Society of the Philippines. Manila.

Tuggy, D. 1993. Ambiguity, polysemy and vagueness. Cognitive Linguistics. 4:273–290.

van Hoek, Karen. 1997. Anaphora and conceptual structure. University of Chicago Press.

Doug Inglis, Box 135, Chiangmai 50000, Thailand. E-mail: doug_inglis@sil.org

BEST COPY AVAILABLE

How the brain evolved language. By DONALD LORITZ. New York/Oxford: Oxford University Press. 1999. 227 pp. Cloth $45.00.

Reviewed by ALAN S. KAYE
California State University, Fullerton

Books on 'language and the brain' seem to be proliferating lately. Even some well–known general theoretical linguists are writing on this topic; e.g. Lamb (1999) for which see Kaye (2000). The well–researched and well–written tome undergoing review begins with the observation that a baby wildebeest born on the Serengeti Plain in Tanzania learns to walk and run in a few minutes, whereas a human infant takes two years to learn to say a word (p. 3). Most of this work attempts to put the miracle of language into proper perspective from an evolutionary viewpoint, using what Loritz refers to as **adaptive** grammar. This theory is based on Stephen Grossberg's (1982) adaptive resonance theory. The author puts it this way (p. 16):

> [It attempts to provide] a theory of how human language, which functions to serve our minute–by–minute social adaptation, arose as part of the same adaptive, evolutionary process which led to Homo loquens.

Let me begin with a comment on Chapter 2, 'Jones's theory of evolution' (pp. 21–35), in my view the most important section for general linguists. Here it is shown that Charles Darwin did not so much come up with the theory of evolution, but rather merely correctly applied it to biology. It was Sir William Jones, in his famous 1786 address to the Royal Asiatick Society of Calcutta (Bengal), who first came up with the idea as applied to language: ' ... indeed, no philologer could examine them all three, without believing them to have sprung from some common source, which, perhaps, no longer exists' (p. 21). Loritz explains that, since there was no Darwinian–type evidence that language evolved, Chomsky (1988) was, according to Loritz, prompted to assert that ' ... any attempt to study language as having evolved from general animal intelligence was "adaptationist hogwash" ' (p. 22). It was, however, research into teaching American Sign Language (ASL) to chimpanzees and apes that led scientists to conclude that not only were speech and hearing wired into the brain, but so too was general intelligence. Philip Lieberman's (1968) findings that apes could not learn to produce human speech[1] due to physiological (i.e. anatomical) limitations led

[1] Philip Lieberman is probably the most qualified researcher in this field. See for further details the outstanding work reported in Lieberman (1975, 1984, and 1991).

researchers such as David Premack to teach them ASL. Many ethologists demonstrated a link between the study of animal and human language, yet most linguists were unconvinced that apes could produce novel utterances. When one of the best-known chimpanzees, Washoe by name, signed 'water' and 'bird' together upon seeing a 'swan', how could one be sure this was a novel utterance? More reasonably, 'water bird' was rather two complete thoughts: 'water' is there, and so too 'bird' is there (p. 22). Human beings were thus still alone as the masters of syntax. As the years went by, even some of Chomsky's MIT colleagues began to abandon his uncompromising position that language did not evolve. Loritz sums up where we are at present by stating (p. 23):

> For a deeper explanation of language, we must look deeper into evolution. We must go back in time, long before the hominids, long before the prehominids. In order to understand how human language and communication make survival possible, we must understand how intercellular communication among the first one–celled life forms made multicelled life possible. And to understand this, we must go back long before dinosaurs, back to when the only living things were rocks.

In other words, language has evolved, and furthermore it is the duty of scientists to account for this evolution as well as the complexity of language. Loritz goes on to deal with the communication of one–celled organisms, which gave rise to multi–celled life forms. Slowly, the reader comes to understand that thought and language can but should not be confused. Language is serial, but it does not necessarily follow that the computer (read: the brain) behind language is a serial processor. Thought is parallel and the brain is a parallel processor because a two–celled brain (at the very minimum) requires a parallel organization, since both cells are simultaneously and concurrently firing their synapses. Chomsky is taken to task again, as the author quotes Chomsky's (1988) assertion (p. 13) that 'no one knows anything about the brain.' Loritz maintains the opposite perspective, with which I am in tirm agreement (p. 14):

> The problem today is not that 'no one knows anything about the brain'. The problem is that we know so much about the brain and its abnormalities in so much detail that it becomes difficult to step back and see how the brain might do something so normal and so large as language.

This is a very technical book for linguists in that linguists, most of whom such as myself, lack the background in physiological psychology. However, some things are (thankfully) quite familiar—such as the fact that syntax is normally much better learned than pronunciation in second language acquisition (p. 192), but even known territory can have new underpinnings. I did not recognize, e.g. that the reason for this is that syntax does not

BEST COPY AVAILABLE

directly involve the cerebellum (ibid.).[2] I learned many other interesting facts and theories while reading this tome; e.g. that the effusion of testosterone triggering sexual development apparently affected brain development in utero, and many left–handers may have been brain–damaged in infancy, which may account for the statistics that boys are more likely to be left–handed, and also language–delayed and dyslexic[3] (pp. 174-175). However, Loritz notes that 'left–handers wind up being, if anything, slightly superior to right–handers in their verbal skills' (p. 175). This is due to the fact that the right hemisphere will take over the functions of the left hemisphere, the only loss being the delay in the translocation of usage to the other hemisphere, which causes the delay in language acquisition.[4]

This volume convinces me that its subject matter is an important area for cooperative interdisciplinary research efforts, which have enormous possibilities for future breakthroughs in fields such as speech and language pathology and communicative disorders.

References

Chomsky, Noam. 1988. Language and politics. Montreal: Black Rose Books.

Grossberg, Stephen. 1982. Studies of mind and brain: Neural principles of learning, perception, development, cognition and motor control. Dordrecht: Reidel.

Kaye, Alan S. 2000. Review of Lamb (1999). Language 76:2:497–8.

Lamb, Sydney M. 1999. Pathways of the brain: The neurocognitive basis of language. Amsterdam: John Benjamins.

Lieberman, Philip. 1968. 'Primate vocalizations and human linguistic ability'. Journal of the acoustical society of America 44:1574–1584.

____. 1975. On the origins of language: An introduction to the evolution of human speech. New York: Macmillan.

____. 1984. The biology and evolution of language. Cambridge: Harvard University Press.

____. 1991. Uniquely human: The evolution of speech, thought, and selfless behavior. Cambridge: Harvard University Press.

Alan S. Kaye, California State University, Fullerton, CA 92634-9480.
E-mail: akaye@Exchange.FULLERTON.EDU

---

[2] One should keep in mind that language is processed in the left cerebral hemisphere, which seems to be related to the fact that most people are right–handed.

[3] When taking my son for speech therapy as a child, I recall wondering why 90 percent of the children in the waiting room were boys.

[4] See also fn. 2.

BEST COPY AVAILABLE

Comparative syntax of Balkan languages. By MARIA–LUISA RIVERO and ANGELA RALLI, eds. Oxford Series in Comparative Syntax 17. New York: Oxford University Press. 2001. 234 pp. Cloth $ 49.95.

Reviewed by JULIE MAY KOLGJINI
The University of Texas at Arlington, TX (Graduate Student)

The purpose of this volume is to describe various syntactic characteristics which are found in the languages of the Balkans. The Balkan languages that are analyzed are: Albanian (Standard), Bulgarian, Greek (various dialects including Modern, Ancient, New Testament, and Medieval), and Romanian. The theoretical slant of the book is based on Chomsky's formalistic minimalist approach. Brian Joseph introduces the volume by discussing the importance and relevance of the languages that are spoken in the Balkan Sprachbund. The topics discussed are as follows: Head–to–head merge in Balkan subjunctives and locality (Carmen Dubrovie–Sorin), Control and raising in and out of subjunctive complements (Anna Roussou), Subjunctives in Bulgarian and Modern Greek (Iliyana Krapova), Direct object clitic doubling in Albanian and Greek (Daliana Kallulli), Adjectival determiners in Albanian and Greek (Antonia Androutsopoulou), and Last resort and V movement in Balkan languages (Maria Luisa Rivero).

Although this volume offers not only a formalistic theoretical approach but also a comparative analysis concerning the languages spoken in the Balkans, it fails to bring to light additional languages (e.g. Serbian and Croatian) and non–standard varieties (e.g. Gheg Albanian) in this linguistic region. Also, although Turkish is usually not considered one of the languages in the Balkan Sprachbund, it has contributed much to many of the languages in this region; given this, one might have expected more mention of Turkish to have been made. Further, if one is not familiar with the minimalist program, the details of the arguments presented in this book could be challenging to understand. Lastly, because of this, it might not be too useful to a field linguist when working with the more 'messy' aspects of language.

Julie May Kolgjini, 9471 Irving Avenue No., Brooklyn Park, MN 55444.
E-mail: juliemay@prontomail.com

Eastern Kayah Li: Grammar, texts, glossary. By DAVID SOLNIT. Honolulu: University of Hawai'i Press. 1997. 385 pp.

Reviewed by KEN MANSON
SIL—MSEA Group, Thailand

Eastern Kayah Li (EKL), or Eastern Red Karen, is a Tibeto–Burman language of the Central Karenic branch. It is most closely related to

BEST COPY AVAILABLE

Bre/Bwe. This grammar is a reworking of Solnit's Ph.D. dissertation (University of California, Berkeley 1986) and is based on 21 months fieldwork in Mae Hong Son. It is interesting to note that it took nine years for this book to be published. In fact, the only references cited in the previous five years before publication were to two conference papers by Solnit and an unpublished manuscript by Fraser Bennett.

The book is primarily a descriptive grammar drawing eclectically from a number of theoretical orientations, not all of which are explicitly referred to. It is divided into three parts as the title suggests: grammar (310 pages), texts (23 pages), and glossary (40 pages).

Looking at Part I, the first chapter is a 'Typological Outline' (eight pages) providing a brief overview of EKL phonology and grammar. The clause structure diagram (p. 6) appears to be a cross between an Aspects and X–bar structure. This structure does not capture anything to do with focus/topicalisation that is discussed later in the book. The paragraph on EKL tone presented states that there are four tones, yet in the following chapter five tones are discussed.

Chapter 2, 'Phonology' (18 pages) describes briefly the syllable structure (standard Tibeto–Burman), the phonemes (initials, clusters, and rhymes), the tonal system, prefixes (phonologically dependent initial syllables of historical importance), and some phonological processes. I would have preferred to find the discussion on the orthography used in this book to have been part of chapter 1.

Chapter 3, 'Morphemes, Word Formation, Grammatical Categories' (27 pages) discusses various word level features of EKL. Morphemes are monosyllabic, words are a combination of full syllables (and prefixes), and the most common word formation process is compounding.

The notoriously difficult verb phrase is taken up in Chapter 4, 'The Verb Complex' (87 pages). It was interesting to note that verb serialisation in EKL follows the pattern for verb final languages rather than for verb medial languages in not allowing any noun phrase to intervene. This chapter draws on a number of insights from Role and Reference Grammar, but unfortunately is based on an early formation of the theory (Foley and Van Valin 1984), rather than more recent and complete versions (Van Valin 1993, Van Valin and LaPolla 1997) that provide a better framework for discussion. More than two–thirds of the chapter look at V+V combinations based on a number of semantic characteristics (causativity, resultative, modal, etc.), a few pages are given to predicates with more than two verbs, and the remainder of the chapter describes verb particles.

BEST COPY AVAILABLE

In chapter 5, 'The Clause' (33 pages), grammatical relations are discussed. The distinction between topic and subject is drawn; there is also a discussion of the object (both direct and indirect) and obliques—non–core roles. It was of great interest to observe that in EKL the classifier phrase (number + classifier) tends to occur clause final rather than with the noun phrase it is modifying.

Chapter 6, 'The Noun Phrase' (46 pages) provides a description of nominal features and order in EKL (p. 179 gives the linear order of constituents). Nearly half of this chapter is taken up with a description of classifiers.

Chapter 7, 'Sentence Types and Sentence Particles' (25 pages) presents a broad catalogue of possible sentence types based on two features—autonomy and nominalisation.

- [- autonomous, - nominalised] eg. headless relative clauses
- [- autonomous, + nominalised] eg. headed relative clauses
- [+ autonomous, - nominalised] eg. basic independent sentences
- [+ autonomous, + nominalised] eg. complement clauses

Sentence particles in EKL include negation, aspect markers, and illocutionary force particles.

'Interclausal Syntax' (23 pages) is tackled in chapter 8. Basically three types of dependent clauses are discussed—relative, complement, and adverbial clauses. It is a shame that RRG theory of interclausal relations was not used as it provides the most explicit theory regarding interclausal relations. This is certainly an area for further research.

Chapter 9 considers 'Elaborate Expressions and Parallelism' (17 pages) providing a number of examples of these four–part constructions.

The final chapter, 'Dialects, Position in Karen, and Orthography' (21 pages) discusses the varieties of Kayah Li, notably the east–west dialect distinction, as well as looking more closely at some varieties of EKL. A short note on the Burmese based script is appended to the chapter.

Part II (23 pages) provides four short glossed texts—two narratives and two conversations.

Part III (40 pages) presents a short listing of all the words used in the texts plus 'selected core vocabulary' (note that 16 words from the Swadesh 100 were missing). There are just under 900 words listed. Also provided was an English–EKL index.

Typographical errors are rare, eg. page xviii, /E/ instead of /e/. I would recommend this book to everyone working in Karenic languages and would also suggest that those working in Tibeto–Burman languages be at least acquainted with the book. Solnit does provide a good well–researched grammar of EKL, and now we must go on from there and fill in the gaps.

REFERENCES

Foley, William A., and Robert D. Van Valin, Jr. 1984. Functional syntax and universal grammar. Cambridge: Cambridge University Press.

Van Valin, Robert D., Jr. 1993. A synopsis of role and reference grammar, 1–164 pp. In Advances in role and reference grammar, Robert D. Van Valin, Jr., ed. Amsterdam: John Benjamins.

Van Valin, Robert D., Jr., and Randy J. LaPolla. 1997. Syntax: Structure, meaning and function. Cambridge: Cambridge University Press.

Ken Manson, Box 317, Chiang Mai 50000, Thailand.
E–mail: ken–roslyn_manson@sil.org

Handbook of natural language processing. By ROBERT DALE, HERMANN MOISL, and HAROLD SOMERS, eds. New York: Marcel Dekker. 2000. 943 pp. $195.00.

Reviewed by MIKE MAXWELL
Linguistic Data Consortium, University of Pennsylvania

The *Handbook of natural language processing* is divided into three sections: 'Symbolic approaches to NLP' (traditional, algorithmic approaches to NLP); 'Empirical approaches to NLP' (statistical approaches); and 'Artificial neural network approaches to NLP'. Within each section, there is an introductory chapter, followed by a half–dozen chapters on specific techniques, and several more chapters about applications using the methodology of that section. Each chapter is written by one or more specialists in that field, while the book editors each edited a section (including writing that section's introductory chapter). A list of the chapter titles would take up the rest of this review, but can be accessed at the publisher's website, http://www.dekker.com. Spoken language technology is not covered.

As one might expect, individual chapters vary in quality, but in general meet the criteria that I would expect of a 'handbook' (but see below concerning the editing standards). Namely, a handbook should be:

(1) up-to-date;
(2) comprehensive;
(3) definitive, but not presenting an overly narrow viewpoint;
(4) well indexed; and
(5) serviceable as a pointer to the primary literature.

The introductory chapters in the first two sections give very brief overviews of their subject material, followed by summaries of the chapters in those sections. Hermann Moisl's introductory chapter for the third section departs from this by delving much more deeply into Artificial Neural Network (ANN) technology, since the editors expected that most readers would be less familiar with that approach. The ANN approach is there contrasted with the symbolic approach (the topic of the first section of the book), with a list of advantages and disadvantages. But although it is mentioned that 'ANNs are now seen as one approach to statistical modeling' (p. 669), ANN technology is not contrasted with traditional statistical approaches (the topic of the second section).

Most of the methodology chapters seem to be similar in terms of their scope and their depth of coverage. Algorithms are described using pseudo–code, equations given where appropriate, and diagrams where helpful. Sufficient information is usually given to understand the pros and cons of individual approaches, if not to actually implement them (for that, one should consult the original references). For example, Dekai Wu's chapter on 'Alignment' presents a history of statistically–based methods which have been used to align bilingual texts, pointing out where individual methods succeed or fail because of language–particular factors. Where alternative approaches cross the boundaries between symbolic and statistical methods, these are also compared; John A. Carroll's chapter on statistical parsing, for example, devotes considerable attention to the relative merits of statistical and rule–based parsing. Unfortunately, in the chapters on ANNs, most of the comparisons are internal to the overall ANN approach. Perhaps this is because the authors of did not feel the ANN methodology was mature enough to merit comparison to more traditional methods. Georg Dorffner comments (p. 789) that 'For engineering applications ... there do not seem to be many advantages of using a connectionist approach to interpret or generate word forms over a rule–based one.'

Occasionally a chapter's coverage goes beyond the bounds of the section within which it appears. Notably, Rajesh G. Parekh and Vasant Honavar's chapter on 'Grammar Inference, Automata Induction, and Language Acquisition' appears in the section on ANNs, but learning by neural networks takes up only a small part of the chapter. The authors devote

considerable space to learning methods from other camps as well, including both algorithmic and statistical approaches. (I hasten to add that I view this broader coverage as positive; it is the placement of the article, not its contents, that seems odd.)

The chapters on applications of the various approaches vary somewhat more than the other chapters in scope and quality. The topic which George E. Heidorn discusses is very narrow: the development of Microsoft's Word 97 grammar checker. Heidorn's discussion even delves into dialog boxes (complete with pictures), right clicking, and 'balloon-style' on–line help. Several other application chapters similarly focus on a particular project, but most are at a broader level. Harold Somers' article, for example, is an excellent summary of a variety of work done by different groups on example–based machine translation. Similarly, David Yarowsky's chapter surveys statistically based approaches to word sense disambiguation.

Chun–Hsien Chen and Vasant Honavar discuss using ANNs for 'noise–tolerant information retrieval and database query processing'. But their discussion is mostly theoretical: they give a 'proposed neural architecture' (to be implemented in special hardware) and compare its 'anticipated performance' with some other approaches. Moreover, it turns out that this application has very little to do with natural language; rather, it is string pattern matching, and the search could just as well have been for DNA sequences or even numbers. The literature on string pattern matching is vast, and there are a number of algorithmic methods that might have fared better than the methods Chen and Honavar actually use as their basis of comparison with ANNs.

The editors state that they have not attempted to 'impose uniformity on the individual parts of the handbook'. I consider this a defect, not a blessing—indeed, the handbook's worst failing is arguably the lack of a strong editorial hand. For starters, there are misspellings (of the kind that should have been caught by a spell checker); typos (ranging from missing punctuation to transposition of entire sentences from one paragraph to another); codes in the middle of text (probably typesetting commands); missing or incomplete dates in references (e.g. '19??'); unparsable sentences; and more.

Graphics are often low resolution (one such low resolution graphic was used on the book's cover, but at a higher resolution). While for the most part the low resolution is simply annoying, occasionally it obscures the meaning. One picture shows un–enhanced and signal processed dot plots; the low resolution view makes the signal processed plot appear miraculous.

BEST COPY AVAILABLE

Figures in most chapters appear in the text about where one would expect; but those of chapter 31 appear together at the end of the chapter. Captions are sometimes incomplete, leaving the reader to puzzle out the meaning of the pictures, or just give up. A graph on page 536, for example, seems quite impressive until one notices that the axes are labeled as 'Plausibility' and 'time', with no numbers on either axis. Equations are sometimes copied from the original papers, but without copying the definitions for all the variables—defeating the purpose of including the equations in the first place.

The lack of uniformity in editing extends to the citation systems, of which there are several. Worse, checking the citations in the text against the lists of references at the end of each chapter seems to have been left up to the individual chapter authors. The result is occasional (or not–so–occasional) missing or incorrect references: in Yarowsky's article, one out of five citations is incorrect or calls for a non–existent reference. This last point, in particular, goes beyond nit picking: a handbook *must* reliably lead the writer to the original papers. When citations are wrong, the handbook fails in that purpose.

Ironically, in one area where the editors have imposed a style, it is less helpful than it could have been. Each chapter uses outline headings, which are certainly helpful. But the outlines would have been even more useful if the headings had also appeared at the beginning of each chapter, as a table of contents. As it is, the outlines help if you are reading the entire chapter, but not if you are searching the chapter for a particular topic.

The index is reasonably complete, but again, it is inconsistent. To take one example, there are entries for 'bottom–up filtering' and 'bottom–up prediction', but not for 'bottom–up parsing': one must instead look for 'parsing, bottom–up.'

I have railed at length about the editing problems of this handbook. Lest I finish with the wrong impression, I want to say that this is a good book; a second (corrected) edition would be even better. At the very least, the publisher should make errata pages available.

Finally, this handbook deserves to be compared with another recently published book, *Survey of the State of the Art in Human Language Technology* (reviewed here in *Notes on Linguistics 3.3*). First, coverage: the *Survey* concentrates on traditional approaches to NLP, with just a few pages concerning neural network approaches and statistical approaches, each of which constitutes an entire section in the *Handbook*. The *Survey*, on the other hand, devotes chapters to spoken language input and to written (handwritten and printed) input, topics accorded just a few pages in this

BEST COPY AVAILABLE

Handbook. (Optical Character Recognition is given a dozen pages in the *Handbook* as an ANN application.)

Second, the *Handbook* goes into details of implementation far more than the *Survey*. Pseudo–code and equations are frequently used in the *Handbook*, but are virtually absent from the *Survey*. As a broad generalization I would say that the *Survey* is well suited as a guide for managers who want to evaluate whether NLP can perform certain tasks (or even the well–educated layman who wants to know about NLP), while the *Handbook* is directed towards implementers of NLP technology.

Thirdly, the *Handbook* is slightly newer, having been published in 2000, while the *Survey* appeared in 1997. (In both cases, the most recent references are several years before the actual publication date.)

And lastly, the price of this *Handbook* will put it out of reach of many, while the *Survey* is available in printed form at a quarter of the price, and on–line for free. This should, however, be balanced against the fact that the *Handbook* is bigger than the *Survey*: over 900 pages of text vs. about half that many somewhat smaller pages.

Mike Maxwell, c/o Linguistic Data Consortium, 3615 Market St, Suite 200, Philadelphia PA 19104. E-mail: maxwell@ldc.upenn.edu

Cambridge history of the English language, IV, 1776-1997. SUZANNE ROMAINE, ed. New York: Cambridge University Press. 1998. 783 pp. Hardback $125.00.

Reviewed by CHARLES A. MORTENSEN
SIL—Colombia Branch

Cambridge history of the English language, Volume IV, 1776–1997 (hereafter *CHEL IV*) is part of a six–volume set edited by Richard M. Hogg. The purpose of the series, as printed on the book jacket, is to 'provide a full account of the history of English. Its authoritative coverage extends from areas of central linguistic interest and concern to more specialized topics such as personal ... and place names.' Volume I, edited by Hogg, is subtitled *The Beginnings to 1066.* Volume II covers 1066 to 1476 and is edited by Norman Blake. Volume V is entitled *English in Britain and overseas: Origins and development* and is edited by Robert W. Burchfield. Volume III, still forthcoming as of the printing of Volume IV, deals with the period from 1476 to 1776 and is edited by Roger Lass. Volume VI, also

forthcoming, is titled *English in North America: Origins and development* and edited by John Algeo. Within Volume IV, the focus of this review, the other volumes of the series are abbreviated as *CHEL* with the appropriate number following.

CHEL IV consists of seven articles which deal with the history of the English language from 1776 to 1997. The year 1776 is chosen as the symbolic date of the beginning of the period because it is the year in which the United States declared its independence from Britain and in so doing began a move toward linguistic independence. It is also a date consistent with the beginning of some significant developments in English syntax and some major changes in literary language and philosophy.

The introductory article by editor Romaine explains both the purpose of *CHEL IV* as well as the 'changing socio–historical setting in which English has developed in response to a continuing background of diversity as it was transplanted to North America and beyond' (book jacket). Romaine points out that the editors had considerable deliberations over the purpose for *CHEL IV*. Whereas the first three tomes dealt with the development of a standard for English in Great Britain (p. 6) ...

> The purpose of this volume is not to discuss the forms of American and British English [but] rather to lay a common historical foundation on which volumes V and VI may build in their discussion of regional forms of English which developed after 1776.

The obvious historical setting, the American Revolution, not only brought about a new political nation but a perceived new language as Webster developed his 'American' dictionary over against Samuel Johnson's 'English' volume. Development of an Australian dictionary is mentioned as well.

Romaine discusses other social changes which have had diversifying or unifying effects on the language, including codification and literacy, urbanization and linguistic stratification and the accompanying shibboleths, radio and television and the resulting contact between diverse groups of society, feminism and political correctness, the spread of the Cockney dialect even to the speech of the royalty, and the influence of American speech and economic power on British English. Romaine even brings up the possibility of mutually unintelligible Englishes, but goes on to argue that speakers of English can agree on common cores in lexicon, grammar, phonology, and literature (pp. 22–47).

Romaine concludes by asking an open–ended question: Is the English language a success story? She presents a few paradoxes as fodder for the debate: English has spread over the whole world but has not yet conquered the languages in its proverbial backyard nor is it the official language in any major Anglophone country (pp. 55–56). Another paradox is that on the one hand English is exported to many foreign countries while on the other hand native speakers even of the same dialect are not in agreement on some issues of pronunciation, grammar, and usage.

The second article (pp. 57–91) is entitled 'Vocabulary' and written by John Algeo, who begins by chronicling major vocabulary studies done since the middle of the 19th century when such research became popular. Algeo stresses that the English lexicon has grown so much as to be 'potentially unlimited in size' (p. 61). The problem is how to actually measure that growth. Algeo points out that the number of words in the *Oxford English Dictionary* (*OED*) is often used as a point of comparison for such measurements. However, Algeo takes exception to the *OED* as such a standard because of its bias 'in favour of literature and particularly of canonically enshrined authors' (p. 63), its lack of American data, and its inaccurate dating of the first uses of some words.

Much of the article is a discussion of how English words have been formed, their meanings changed, and how they have been shortened, blended, or borrowed. Some statistical evidence is given to demonstrate what percentages of today's words fit into the above categories (blends, borrowings, etc.). One parenthetical section (2.8) shows the languages from which today's English words originate. The final section presents very recently developed vocabulary and the processes by which they have come into the language. Culture change is cited as a major reason for such neologisms. 'Vocabulary, more than any other aspect of language, is inextricably connected with our total culture' (p. 91).

The third article is written by David Denison and is entitled 'Syntax'. This chapter reveals the most revealing and surprising information about changes in English of any of the articles in *CHEL IV*; while new vocabulary items, for example, are noted by the average speaker of the language, Denison points out striking examples of changes in syntax as well as many more subtle ones. He comments on the noun phrase, the verbal group, other elements of the clause, clause structure, and composite sentences. (He declines to discuss morphology at length because the language has not demonstrated appreciable change in that regard.) That this is the longest article (pp. 92–329) in the volume owes itself to the many syntactic changes Denison presents and the wealth of examples from many literary sources by

which he documents them. The points are not overdone and the whole article is well worth reading.

'Onomastics' is the title of the fourth chapter. Author Richard Coates discusses changes in individual names as well as developments in naming strategies. Coates not only treats the subject of personal names but also of surnames, towns, geographical features in the U.S., and streets. While this is not a lengthy article (pp. 330–372), it invites much commentary.

The section on personal names reviews naming innovations prior to 1776 and chronicles trends since that year, not only in which name a child is given but how many names one has and how one might change his name. Regional varieties of given names are also discussed. The discussion of surnames is brief because there have been few additions to the English–based surname stock in the period under study; spelling and pronunciation changes are commonplace, however, whether because of error or vogue. This section ends with some observations on hyphenated surnames or those combined otherwise.

Coates cites several reasons for development of place names. The first, of course, is the settlement of America—more places needed to be named. Secondly, Americans introduced new naming strategies such as numbering of streets, which were laid out in grids, and the use of geographical features to name towns. The author also discusses reasons for changes in spelling of town names. Other name changes were due to rival names for the same place or confusion between two similarly named places. Names were also borrowed from other areas or countries and also from foreign languages. These names were subject to pronunciation and spelling corruptions galore, whether they came from prestigious European languages or North American indigenous languages. Coates cites the name *Tioga,* Iroquois for 'at the forks' but commonly thought to mean 'gate' and widely used to name places which are gateways in one sense or other. One hotbed of such corruption is California, with its morphologically incorrect Spanish names, but the author does not mention these.

Coates also points out that compound place names indicate the center of a rural area in [New] England (e.g. Greenwich Village). However, he omits any mention of the same construction being used to name a prestigious and sometimes closed community in the rest of the U.S.

Coates continues with a section on street names. He chronicles the development of these from the fairly narrow choice of historic persons and events at the beginning of the period under study to the almost unlimited choices city planners have today. He mentions the use of backward

BEST COPY AVAILABLE

spellings, acronyms which become putative words, numbered streets in the U.S., and the use of generics other than STREET, like ROAD, WAY, etc. The U.S. has a much greater variety of choices, he claims; one that is mentioned is CLOSE, a common British generic, which I find means *cul-de-sac*; I am inclined to think it is pronounced with [z] rather than [s].

The article closes with a number of references to studies of a narrower scope, such as the naming of fundamentalist churches in America.

The fifth article in *CHEL IV* is by Michael K.C. McMahon and entitled 'Phonology'. McMahon begins with the disclaimer that the vowels and consonants of today's English were largely in place by 1800. The intent of his article, then, is to demonstrate how pronunciation of certain words and certain phonemes has changed. The evidence is that British English has changed more than North American has.

Piecing together the clues of how English was pronounced 200 years ago is not easy yet McMahon ably manages to decipher the transcriptions of the orthoepists, the studies of whom were the first attempts at analyzing English phonology. Much of their commentary is dedicated to prescribing pronunciation, usually their own. McMahon points out that even then there was debate on how to pronounce certain words; different general accents of English in Great Britain and in North America are rooted in the variant pronunciations. Some of these are different low vowels, /h/ vs. no /h/, intrusive *r*, and the presence or absence of /l/ before /f/ and /k/. The author gives perhaps too much space to discussing phonetic realizations of unstressed vowels but, to his credit, he presents copious amounts of data.

Much space (pp. 492–517) is also given to displaying changes in stress placement in multisyllabic words.

The sixth chapter is entitled 'English Grammar and Usage' and is authored by Edward Finegan. Contrary to what one might assume from the title, this article does not chronicle changes in grammar rules as does Denison's 'Syntax'. It is an account of the difficulties of the standardization of English and the debate over whether correct grammar can even be taught. Finegan stresses that the main difference from the beginning of the period until the present is the great amount of data available for study, analysis, and decision making. The following statement by Finegan more than adequately summarizes this point (p. 586):

> During the 18th century, when grammar started to emerge from philosophy and when the influence of Latin waned, grammarians analysed English because they perceived it to be in need of codification as compared to the classical tongue and

> the Continental vernaculars [pp. 536–40]. During the nineteenth century, however, fanciful etymologies were enlisted in support of inquiry into philosophical truth [pp. 554–73]. When Dean Trench transmuted Horne Tooke's philosophical etymologies into theological inquiries, he provided a compelling basis for viewing linguistic expression as fossilized ethics, a veritable window on past morality and immorality alike. It was then only a short step to viewing contemporary usage through ethical lenses and trying to shape it to what were imagined its best and most ethical forms. The nineteenth century was an era of moral and ethical philology, and ironically Dean Trench, the prime propagator of such a perspective, was also the catalyst responsible for instigating the Philosophical Society's *New English Dictionary* [pp. 559–64]. Once published, the *OED*, with its massive record of facts about English usage, forever made the ethical analysis of grammar and lexicon difficult, though the ethical analysis of language use (quite a different matter) remains strong at the end of the twentieth century.

The immediately preceding comment by Finegan lauding the publication of the *OED* is balanced in *CHEL IV* by Algeo (p. 63) who decries it as being rather unrepresentative of the English language. In spite of the now evident standardization of English, such differences of opinion existed at the beginning of the period under study and continue to this day. That is Finegan's other point!

The seventh and final chapter is by Sylvia Adamson and is entitled Literary Language. Adamson attempts to reconcile the following two commentaries on Modernism in poetry. She quotes on p. 589 the Modernist T.S. Eliot who wrote in 1942 (*The Waste Land)*, 'Every revolution in poetry is apt to be a return to common speech.' The anti–Modernist, Larkin, is quoted on p. 590 as having commented in 1983: '... up to this century literature used language in the way we all use it ... The innovation of "modernism" in the arts consisted of doing just the opposite'.

Adamson ably shows how a device like parenthesis (pp. 592–596) went from simply representing conversational speech (the goal of Romantic poetry, from the beginning of the period under study) to becoming stylized, the ideal of much Modernist verse. One reason for this is that instead of introducing new standards, Romanticism simply relaxed or removed specialized Classical standards to the point at which there was no unitary literary form. The demise of one such standard, heroic verse, is discussed in section 7.3, 'Breaking the pentameter', in which the author not only shows that heroic verse was abandoned but that several other meters, including unnatural triple meter, were employed in harsh reaction to it (p. 614–620). Other devices are mentioned, all of which contribute to the development of blank verse in this century.

BEST COPY AVAILABLE

In section 7.4 (pp. 630–645) an analogous change in prose is chronicled: the loss of highly specific subordinating conjunctions to less specific coordinating conjunctions to mere juxtaposition of clauses. The result of this process, which Adamson calls the information deficit (p. 639), is that the reader cannot always be sure of the relationship between two clauses.

The author also chronicles how metaphor (section 7.5) was criticized in the late Classical period by philosophers, literary critics, and scientists alike as thinly veiled falsehood. Wordsworth resurrected metaphor but allowed it to be checked (as it has been ever since) by 'empirical content and descriptive precision' (p. 647). Some of these 'acceptable' forms of metaphor are chronicled by the author. Adamson summarizes this section by citing Davie (p. 661), who argued that 'the *restraint* of metaphor is the key to achieving a style which maintains both the continuities of the literary tradition and the common–ground between literature and "the very language of men" '.

Romantic literature also reflects the personal or internal realm rather than the external world. Affective language became much more common; Adamson on p. 665 cites Jane Austen as an example of one who not only propagated the use of these but the subjectivization of them. The development of 'subjective' tenses (present progressive and present perfect; see also Denison, pp. 130ff), the conversation poem, the dramatic monologue, and empathetic narrative all invited readers to know characters intimately. Conversely, the author juxtaposes this with the impossibility of it in Modernism (pp. 674ff) and shows how various Modernist writers have managed to dispense with Romanticist elements. She also cites the New Journalists, who attempted to bridge the gap between the two (pp. 677–679).

In summary, Adamson admits that she has presented the two philosophies as successive events. However, she asks the question (which she seems to answer herself): are these successive events or competing trends? She cites romanticists who have made their work more empirical and notes that by the 1930s many writers were taking their stylistic models from 'non–Modernist moderns'. While both philosophies attempt to return to the common core of speech, just what is the real common core? (This 'common core' is what editor Romaine assumes in Chapter 1.) Adamson believes and demonstrates that elements of both Romanticism and Modernism are evident in the purveyors of both schools. This downgrades, in the opinion of the author, the shock one can generate by switching styles. Interestingly enough though, neither of these principle styles enjoys the obvious place of Haiku poetry in the curriculum of schools.

Adamson's article concludes with a listing of citations and a list of the cited authors with their years of birth and death.

CHEL IV concludes with a glossary of linguistic terms, assuming that those other than linguists will enjoy the volume. It also serves to clarify what each author means by a given term. Following this is a biography for each article, a general index, and the contents of the other *CHEL* volumes.

I found *CHEL IV,* although voluminous, to be fairly easy reading although at times, not being a lover of poetry, I found that Adamson's article required more careful scrutiny. Nevertheless, nearly all of the material is quite clearly presented, well analyzed, and brings out aspects of English (some fascinating) that could easily be overlooked since the language has not changed nearly as much in the current period as it did in the previous seven hundred years.

Charles A. Mortensen, Dept. PTY 7982, P. O. Box 025207, Miami, FL 33102–5207
E-mail: chaz_mortensen@sil.org

Bantu historical linguistics: Theoretical and empirical perspectives. By JEAN–MARIE HOMBERT and LARRY M. HYMAN, eds. Stanford, CA: CSLI Publications. 1999. 596 pp. Hardback $74.95, paper $29.95.

Reviewed by STEVE NICOLLE
SIL—BTL (East Africa)

This volume, which arose out of a Table Ronde organized by the editors in 1996, is a fascinating book of a generally very high standard. It seriously challenges the view that historical linguistics is irrelevant to field linguists, and conversely demonstrates the value of detailed field–work to historical linguistics. The chapters are grouped under classification, phonology, or morphology, which I will describe in turn.

The first section, 'Classification and linguistic history', begins with chapters by Nurse: 'Towards a historical classification of East African Bantu languages' (1–41), and Ehret: 'Subclassifying Bantu: the evidence of stem morpheme innovations' (43–147). Ehret's use of stem morpheme innovations is well suited to wider grouping, while Nurse's phonological and morphophonological criteria probably provide a more robust basis for narrower grouping. Nurse does, however, provide a tentative analysis of larger groupings which suggests a history of 'constant divergence and convergence' (31) rather than the traditional divergent development easily represented by tree diagrams.

Ehret identifies a Savanna–Bantu group which covers three quarters of the Bantu–speaking area (excluding Zones A–C), which he subdivides into Western–Savanna, with six sub–groups, and Eastern–Savanna, with five sub–groups. Nurse restricts his study to East Africa, corresponding to most of Ehret's Mashariki sub–group of Eastern–Savanna[5]

Despite their different methodologies, Nurse and Ehret arrive at remarkably similar divisions. Both identify the P.30 languages as a group (Nurse's North Mozambique, Ehret's Makua) although Ehret classifies Makua under Mashariki–Kusi. Ehret's Mashariki–Kaskazi (northern) grouping contains nine groupings to Nurse's eleven. They agree on North–East Coast (E.70 minus E.74a Dawida, and G.10–40), Great Lakes (J including D.40–60 and E.10–40), Langi and Mbugwe (F.33–34), and Takama/Western Tanzania (rest of F). Ehret combines Nurse's Central Kenya (E.50) and Kilimanjaro–Taita (E.60 and E.74) into a single group called Upland, and his Mwika–Rungwe equals Nurse's Nyakyusa (M.30) and SW Tanzania (M.10, M.20). Nurse's Southern Tanzania Highlands (G.60 plus Manda N.11) approximates Ehret's Njombe (G.60 only), and they both identify Kilombero (G.50, but Ehret includes P.15), and Rufiji–Ruvuma (N.10, P.10, P.20, but Nurse omits N.11 and Ehret omits P.15).

Both Nurse and Ehret argue persuasively against relying on lexicostatistics alone, which rather diminishes the appeal of the two lexicostatistically oriented chapters by Bastin and Piron: 'Classifications lexicostatistiques: *bantou, bantou et bantoïde'* (149–163) and Mann: 'A note on historical and geographical relations among the Bantu languages' (165–171). Such arguments do not, however, detract from Hinnebusch's application of lexicostatistics to the identification of contact situations: 'Contact and lexicostatistics in comparative Bantu studies' (173–203). Hinnebusch assumes that 'LS similarities measure genetic and non–genetic variables alike' (178), hence when one member of an otherwise cohesive genetic group diverges lexicostatistically, this demonstrates the operation of non–genetic variables, such as geographical dialect chain relationships, retention of lexis that other members of the group have replaced, or sociocultural influences. The latter is exemplified by Digo (E.73) which has a high proportion of Swahili loans compared with the rest of the Mijikenda group. However, while it is correct that the Digo are more culturally assimilated than other Mijikenda groups to Swahili, eight of the eleven illustrative Mijikenda forms supposedly replaced by Swahili loans in Digo are attested

[5] Nurse does not discuss three out of Ehret's four Mashariki–Kusi (southern) sub–groups: Nyasi (N.20–40 minus N.41), Shona (S.10), and Southeast–Bantu (S.20–60 and K.21).

in the Kenyan Digo that I am acquainted with (although five of these alternate synonymously with the corresponding Swahili loan, one has a modified meaning and one is archaic).

The second section, 'Comparative and Historical Phonology', consists of six chapters. I shall briefly list these (with a few brief comments) but will discuss only the chapter by Hyman in any detail.

John Stewart: 'Nasal vowel creation without nasal consonant deletion, and the eventual loss of nasal vowels thus created: the pre–Bantu case' (207–233) uses comparison across Bantu and non–Bantu languages (primarily Akan and Ebrié) to reconstruct Proto–Bantu phonology.

Denis Creissels: 'Remarks on the sound correspondences between Proto–Bantu and Tswana (S.31), with particular attention to problems involving *j (or *y), *į and sequences *NC' (297–334).

Catherine Labroussi: 'Vowel systems and spirantization in Southwest Tanzania' (335–377) looks at spirantization and the change from 7V to 5V systems. Spirantization occurs when a C was followed by one of the Proto Bantu super close vowels *į or *ụ, (either within a morpheme or in a suffix) as in the following Digo examples:

–iba (steal) > *mwivi/mwizi* (thief);
–vyala (give birth) > *avyazi* (parents);
**–dzogụ,* > *ndzovu* (elephant).

Typically, spirantization is more limited in 7V languages than in 5V languages, and is presumed to precede the 7V > 5V merger, but Labroussi describes cases of limited spirantization in 5V languages and apparently full spirantization in 7V languages. To account for the variation she proposes the following diachronic stages:

1) spirantization occurs morpheme internally;
2) 7V > 5V merger may occur;
3) spirantization occurs across morpheme boundaries in the most innovative languages;
4) necessary 7V > 5V merger occurs (367). If 7V > 5V merger occurs at stage 2 it removes the necessary phonetic environment for further spirantization (374), hence there are 5V languages with limited spirantization.

Thilo Schadeberg: 'Katupha's Law in Makhuwa' (379–394) discusses deaspiration in Makhuwa (Katupha 1983) which applies to all but the last aspirated C in a stem (383), contrasting this with deaspiration/aspiration shift in Swahili which causes only the first suitable C to be aspirated. Finally, he

BEST COPY AVAILABLE

proposes the following relationship between Katupha's Law and Dahl's Law:

> (a) At some stage all voiceless stops became aspirated: [p^h, t^h, k^h].
> (b) Katupha's Law reduced the first in a sequence of two aspirates to a plain stop.
> (c) Dahl's Law changed unaspirated voiceless stops into voiced ones (392).

Gérard Philippson: '*HH and *HL tone patterns in the Bemba tone system' (395–409). This has a different title in the contents page: 'Unresolved puzzles in Bantu historical tonology'.

Hyman's chapter 'The historical interpretation of vowel harmony in Bantu' (235–295), stands out for the controversial proposals he makes and the range of data he cites; it is really (at least) two papers in one. To begin, he uses information from 134 languages stored in the Comparative Bantu On–Line Dictionary (CBOLD) project to show how asymmetries in the operation of vowel height harmony (VHH) in front and back vowels reinforces Ehret's identification of a Savanna–Bantu genetic group. He then argues, against most previous accounts, that VHH within verb–stem suffixes developed after Proto–Bantu (PB), and that VHH asymmetry was a Savanna–Bantu innovation.

In asymmetric VHH the rules for front and back vowel harmony are as in (1) (237):

(1) a. front height harmony: *i > e / { e, o } C _
b. back height harmony: *u > o / o C _

i.e., VHH fails to obtain in (*eCu). Asymmetry obtains in almost all 5V languages and some 7V languages, in an area corresponding closely to Ehret's Savanna–Bantu. Within this area, /a/ conditions front VHH in most of zones K and R (242), i.e.:

(2) front height harmony: *i > e / { e, o, a } C _

Front VHH can also be absent, as in Southern Bantu, without loss of back VHH (245), which along with the situation described in (2) provides supporting evidence for the independence of front and back VHH presupposed by (1). (Interestingly, VHH of final vowel and prefixes only occurs in languages with symmetrical VHH and, typically, seven vowels [244].)

Most previous studies reconstruct asymmetric VHH (with seven vowels) for PB (255), but Hyman argues that since the languages with asymmetrical

VHH systems are a more homogenous group than those with symmetrical systems (zones A, B, C and EJ.40), it is more likely that asymmetrical VHH was an innovation. Asymmetrical VHH would then be 'a reliable genetic marker' of Savanna–Bantu (256). Hyman then proceeds to tell a short story with a twist in the tale. He first provides arguments in favor of the view that PB monomorphemic forms (e.g. bisyllabic noun stems) exhibited symmetric VHH, and supports this with reference to languages such as Nande (DJ.42) which exhibit asymmetric VHH in verb–stem suffixes but show symmetric VHH in bisyllabic noun stems. He then shows that this situation is more likely due to independent variation than to inheritance (258–261). His conclusion, then, is that although asymmetrical VHH was an innovation, PB probably did not have a symmetrical system.

This obviously calls for an alternative hypothesis. Hyman 'tentatively' suggests that rather than suppose that all PB verb–stem suffixes had the forms **-iC-* and **-uC-*, we should consider the possibility that some of the verb–stem suffixes had the forms **-eC-* and **-oC-*. This would mean that the rules in (1) would be restated as (3):

(3) a. front height harmony: *e > i / { j, ụ, i, u, a } C _
b. back height harmony: *o > u / { j, ụ, i, u, a } C _

In this case, the motivation for the changes is 'peripheralization', i.e. the tendency for vowels to migrate to the peripheries of the vowel space, becoming /i, u, a/ in 'weak' positions (e.g. internally and/or finally) unless **reinforced**, in the case of *e by a preceding vowel of the same height and in the case of *o by a preceding identical vowel (269). The asymmetry between front and back VHH can be accounted for if we reconstruct some suffixes with degree 3 vowels (e.g. the applicative and stative as **-ed-* and **-ek-* respectively) and others with degree 2 vowels (e.g. causative as **-ic-j-*, reversive intr. as **-uk-* and reversive tr. as **-ud-*). Front VHH would then follow the rule in (3a) (i.e. raising through **peripheralization** unless reinforced by a preceding degree 3 vowel) while back VHH would follow the lowering rule in (1b). The apparent extension of front VHH following /a/ in zones K and R (see above) can therefore be viewed as original; the preceding /a/ also hindered peripheralization (274).

Hyman applies this hypothesis to the solution of a number of puzzles pertaining to individual languages, and then develops the logical conclusion that follows from it: that PB had no stem–internal VHH (276). This claim is supported with evidence primarily from Ndebele (S.44). He concludes by outlining a number of consequences of his revised view of PB.

The third section, 'Comparative and Historical Morphology', consists of five chapters; these are, briefly, as follows. Claire Grégoire and Badouin Janssens: *'L'augment en bantou du nord–ouest'* (413–429) discusses the distribution of nominal pre–prefixes in zones A, B, and C.

Pascale Hadermann: *'Les formes nomino–verbales de classes 5 et 15 dans les langues bantoues du Nord–Ouest'* (431–471) addresses the question of which contemporary form(s) originated from a (PB) infinitive and which from a deverbative. Having observed that in languages with two nomino–verbal forms, the occurrence of each is typically functionally arbitrary or due to dialectal variation, Hadermann looks at metatony (where the final V of a non–finite form takes a high tone if followed by a complement), exemplified here from Buyi (D.55):

(4)	a.	*okáboa*	to build
	b.	*okábóá sɛsɛɛ*	to build a house (441)

Languages with high tone metatony and only a class 15 infinitive apply metatony systematically to this infinitive (445), while languages with two infinitives never have metatonic H with class 5 and only some allow metatonic H with class 15 (448). Hadermann explains this as reflecting the dual verbal and nominal nature of class 15 infinitives; when followed by a complement, the verbal nature predominates and metatony underlines the relation between the infinitive and its complement. Conversely, metatony is blocked with nomino–verbal forms in class 5 because these derive from deverbatives and hence do not enter fully into the verbal paradigm (453). Hadermann concludes by arguing that most languages without a class 15 infinitive show evidence of such a form at an earlier stage in their TAM systems, which supports a class 15 infinitive for Proto Bantu. There was probably a class 5 nomino–verbal form also, but the evidence for this is less convincing.

Robert Botne: 'Future and distal *-ka-*'s: Proto–Bantu or nascent form(s)?' (473–515) looks at the distribution of the TAM morpheme *-ka-* (or *-Vka-*) in its future and distal uses ('go and V', 'V elsewhere'). The distal form is concentrated in the south–central Bantu area, but Botne is incorrect when he says that 'distal markers of any type apparently do not occur at all in zones B, D, J, and S, and only in a few languages in zones A, C, E, F, G, and H' (476). Distal forms occur in Fuliiru (zone J) (van Otterloo 2000), and three of the four zone E languages Botne cites have distal *-ka-*, to which we can add Digo (E.73). The occurrence of distal *-ka-* in infinitives or indicatives almost guarantees its occurrence in imperatives (and to a lesser extent subjunctives) (482), and so Botne concludes that distal *-ka-* may have

originated in imperative constructions, most probably as a locative clitic following verbs of motion which, once deleted, left *-ka-* with the sense 'go and V' (488). Since future *-ka-* invariably occurs in the indicative if it can occur in the infinitive or subjunctive, Botne suggests a separate origin in indicative constructions (496).

Botne claims that future *-ka-* was already grammaticalized in Proto Bantu (504), but this is contradicted by Derek Nurse and Henry Muzale in the following chapter: 'Tense and aspect in Great Lakes[6] Bantu languages' (517–544). Nurse and Muzale argue that the Great Lakes future form *-aka-...-e* is a recent innovation rather than a PB or even Proto–GL reflex, largely on the basis of its limited geographical distribution. However, the tendency for future tenses to be marked by innovations could equally mean that the limited distribution of *-aka-...-e* is due to its having been displaced by more recent innovations. The picture is further muddied for both Botne and Nurse and Muzale by the common occurrence of *-ka-* and *-aka-* forms as past tense markers. In fact Nurse and Muzale propose *-ka-* as a candidate for the Far Past marker in Proto–GL (532–3).

Finally Tom Güldemann: 'The genesis of verbal negation in Bantu and its dependency on functional features of clause types' (545–587) applies grammaticalization theory and diachronic typology to morphological variation in verbal negation. Unfortunately Güldemann's presentation obscures an already complicated picture. The book concludes with a language index of approximately 470 entries.

REFERENCES

Mateus Muária. 1983. A preliminary description of sentence structure in the e–Sáaka dialect of eMákhuwa. M.Phil. thesis, School of Oriental and African Studies, London.

van Otterloo, Roger. 2000. Kifuliiru aspects on stage. Paper presented at the Bantu Initiative Katupha, José symposium, Ruiru (Kenya), February 2000.

Steve Nicolle, BTL (East Africa) and SIL International Digo Language and Literacy Project, P.O. Box 12, KWALE (via Mombasa), Kenya. E–mail: Steve–Alison_Nicolle@sil.org

[6] Title page reads 'lacustrine'

The Athabaskan languages: Perspectives on a Native American language family. By THEODORE B. FERNALD and PAUL R. PLATERO, eds. Oxford Studies in Anthropological Linguistics 24. New York/Oxford: Oxford University Press. 2000. 332 pp. Hardback $55.

Reviewed by Gillian Story
SIL—North America Branch

This book is a collection of articles from a conference at Swarthmore College, Pennsylvania, where the two editors were on teaching staff in 1996. The majority of the articles are concerned with Navajo, an Athabaskan language of Arizona, and one of the editors, Platero, is himself a native speaker of Navajo. Platero and three other Navajo linguists present at the conference hold PhDs in linguistics.

Attendees at the conference included both linguists and educators although the book is almost entirely linguistic in its content. Apparently there were, , discussions which were more including of the educators in respect of their interests and of the ways in which they might contribute to the collection of Navajo linguistic data. The one discussion that is reported in the book concerns the separation of secular and religious elements in the teaching of Navajo culture in public schools. Some argue that this is not possible since traditionally there was no such separation; others either argue that the secular aspects of the culture (including the language) can be taught alone or that Navajo culture can be taught as comparative religion. However, because various faiths are represented among the Navajo people, parental feelings are sometimes against this latter type of approach.

One of the less technical articles concerns particles in Navajo. This is a reprint of a booklet that was published in 1948 and was designed as a help to teachers of the period who were often frustrated in their efforts to teach English to Navajo students. The booklet attempted to show how the function of English intonational patterns in modifying meaning is carried in Navajo by certain structures (mostly grammatical, rarely phonological) and by the use of particles.

Leer's article is the only article which is historical in its orientation and which extends its discussion to members of the Na–Dene language family outside the Athabaskan group. The article is concerned with the negative–irrealis category in Athabaskan–Eyak–Tlingit in its various manifestations. It is a historical survey of the category rather than a presentation of the reconstructive methodology.

Two of the articles are concerned with word order in a discourse context in two dialects of Apache, a southern Athabaskan language (closely related to Navajo), and in Koyukon, an Alaskan Athabaskan language. Both are text–based studies. Jung, in writing about Apache, assumes that it is a polysynthetic language having NP adjuncts rather than arguments. Given that, a pragmatically–controlled word order might be expected, but Jung finds that any violation of SOV order is rare and the out–of–order constituent is outside the core clause (except in cases of voice–related Subject Object Inversion).

Thompson, in writing about Koyukon, is concerned with post–verbal elements in the main with NPs. Preverbal NPs are very much more frequent than post–verbal. Though post–verbal NPs appear not to be 'afterthoughts', Thompson finds their status as integral parts of the clause to be ambiguous. 100 preverbal NPs and 100 post–verbal NPs from continuous text are studied for such factors as syntactic role, human versus non–human reference, definiteness, 'referential distance' from previous mention, and 'persistence' (recurrence of mention in following clauses) (the latter two using Givón's methodology). Post–verbal NPs are generally definite human subjects, and of lower referential distance and higher persistence than preverbal NPs. Post–verbal NPs occur with about the same frequency as preverbal in paragraph initial, medial, and final position. Thompson goes on to consider these results in the light of iconic and cognitive principles. His conclusion is that post–posing old information foregrounds new, preverbal, and verbal information.

Discourse considerations are also the concern of Axelrod in the final section of her article. In the first part, she discusses noun classification in Koyukon (manifested in the verb in classificatory stems and in 'gender' prefixes) and classification of verbs by 'aspectual verb theme categories'. These latter may be viewed as the Koyukon parallel to verb situation types. Similar features distinguish both noun classes and verb theme categories. Both are exploited in discourse to provide cohesion. In the case of verbs, there may be repetition of a verb stem in a different aspect but more interestingly repetition of different verbs in the same aspect referring to the same event.

Three articles are concerned with the question as to whether Navajo NPs are adjuncts or arguments. If there are no PPPs (pure predicate phrases) (of whatever word class) in Navajo, then what typological consequences follow? Faltz' discussion of this question ranges over nouns, determiners (including quantifiers and possessives), relative clauses, postpositions, and attributive adjectives. Willie and Jelinek argue that the ordering of Navajo nominals in the clause is independent of grammatical relations; rather, the ordering is

BEST COPY AVAILABLE

Topic–Focus, not Subject–Object in the direct construction and Object–Subject in the inverse. (Navajo Subject Object Inversion is a misnomer.) The *bi–* prefix that occurs in the inverse construction marks a topicalized patient. Certain other features of nominal syntax are explained by the analysis in terms of topic and focus.

Hale and Platero take as their starting point evidence for movement out of certain negative structures. Movement out of certain NPs in this environment is not possible and since Binding Theory principles indicate that such extraction should not be possible if the NPs are adjuncts, this is possible evidence that Navajo NPs are adjuncts. However, they also find (1) evidence of movement out of possessive NPs suggesting that at least some Navajo NPs may be arguments, and (2) if subordinate structures which are 'GF–parallel' (the subordinate clause sharing arguments of the same grammatical function with the main clause—a feature of Navajo) are arguments, then, if GF–parallelism overrides certain Binding Theory principles, this also points in the same direction. As the authors observe, the evidence is inconclusive.

Rice's article is also in the tradition of generative grammar. She finds that in Ahtna, Slave, and Navajo syntactic properties of a 'monadic' verb depend on semantic factors such as saliency, agentivity, and humanness rather than on an unergative/unaccusative classification.

Two articles discuss aspects of the Navajo verb. McDonough's article is prefaced by two quotes—one of them from Sapir. Speaking of the verb in Haida (whose relationship to Na–Dene is actually a matter of dispute), Sapir says, 'It all crumbles into pieces at the least touch. I think the same will prove to be true of Athabaskan–Tlingit.' McDonough breaks the Athabaskan verb apart into a bipartite structure containing two stems—an inflectional stem and a verb stem, both prefixed. The inflectional stem comprises the conventional tense–mode and subject prefixes in a type of word–paradigm structure. Although McDonough challenges a position class analysis of the Athabaskan verb, her bipartite analysis is phonological in its orientation and does not reflect the hierarchical morphological structure of the verb.

The aim of Smith's article is to characterize the event structure of the Navajo verb in representations that show how different types of information are distributed among the prefixes and the verb stem. These types are discussed in terms of the general categories Figure, Motion, Ground, Path, Cause, Manner, and State. Whereas the typical English verb root expresses a type of Motion and its Manner/Cause, Navajo verb roots typically code a type of Motion and a Figure (the third type of typical verbal lexicalization patterns

BEST COPY AVAILABLE

code a type of Motion and its Path). In Navajo, Ground, Path, and Manner are typically found in the verb prefixes.

Fernald discusses ILPs (individual–level predicates) and SLPs (stage–level predicates) in relation to the well–formedness of generalizations in Navajo. ILPs denote properties that are not tied to location in space and time (e.g. be wooden, be tall, know French); SLPs describe individuals that are so located (e.g. be available, be sick, speak French). The discussion centers on the example 'When Mary speaks Navajo, men respect her'. 'Respect' is an ILP; nevertheless, the sentence is an acceptable generalization because there is a contextual restriction understood—only those men respect Mary who are present when and where she speaks Navajo. Interestingly, generalizations in Navajo are formed according to the same principles of logic as in English.

Educators at the conference probably did not find themselves much the wiser in many of the presentations at the conference and field linguists are likely to find themselves similarly situated.

Gillian Story, 447 Huntsville Cres. N.W. #2, Calgary AB T2K 4W3, Canada.
E-mail: Gillian_Story@sil.org

Mandarin Chinese: An Introduction. By MOBO C. F. GAO. Oxford: Oxford University Press. 2000. 226 pp. Paperback, $19.95.

Reviewed by Terrance R. Wardlaw, Jr.
SIL—East Asia Group

What is the purpose of Gao's introduction to the Chinese language? He states in his preface that this book 'aims to provide an overview of the Chinese language from the perspective of the undergraduate English–speaking learner of Mandarin Chinese' (x). He intends this book to function as a complementary text rather than a primary text for beginning students of Chinese, and as such this book proves helpful as an auxiliary background text for undergraduate college students unacquainted with Chinese culture, geography, or history. Furthermore, he draws all of his examples and analogies from English in order to enable English–speaking students to understand the concepts under discussion.

Chapter 1, 'Language and Setting'(1–28), includes sections that describe China's geography, history, society, rural and urban peoples, language varieties, and the Chinese language family. In this section Gao differentiates the main dialects of Chinese, and then discusses their relationship with the

BEST COPY AVAILABLE

writing system. The beginning student will find this information very helpful as an orientation before digging into the various positions and arguments regarding dialect classification within China.

Chapter 2, 'Language and Politics' (29–51), describes the effects of the past century's political turmoil on Mandarin Chinese. It discusses the differences between spoken and written Chinese, language change in the context of the political environment, language reform in the context of those desiring transformation of China's culture, the influence of anti–traditionalists on language reform, the movement advocating Romanization of the Chinese script, phonetic spelling, and the effects of Communist political movements on the language.

Chapter 3, 'Sounds and Tones' (52–73), includes sections on *pinyin*, syllables, phonetic symbols in Taiwan, various systems of Romanization, tones, loan words, stress, and rhythm. Although beginning texts contain discussions of these issues, Gao's discussion provides a fuller treatment than most.

Chapter 4, 'Writing' (74–94), discusses Chinese characters, radicals, types of characters, writing styles and calligraphy, complex and simplified characters, and characters in countries other than China and Taiwan. This chapter provides much background information and discussion regarding characters, however it is in no way intended to help students learn characters.

Chapter 5, 'Vocabulary' (95–130), treats words, word classes, word formation, compound words, loan words, reduplication, affixes, set phrases, homophones, differences between informal and formal vocabulary, meaning and translation, names, writing addresses and dates, and dictionaries. The section on dictionaries would have been more helpful had Gao included illustrations in addition to his prose description of Chinese dictionaries. This section would also prove more helpful with step–by–step instructions for looking up an unfamiliar character using the tools in the front of a dictionary. Although Gao does not intend to teach students how to use a dictionary with his description, his stated purpose of this text as a complementary textbook suggests that he should have included this type of complementary information for beginning students in order to fully augment the main textbook.

Chapter 6, 'Grammar' (131–180), discusses word order, phrases, question sentences, word order of passive sentences, inflection, grammatical words, measure words, auxiliaries, complements, a summary of the three *des*, quantifiers, and sentences. First–year students will benefit greatly from Gao's discussion.

BEST COPY AVAILABLE

Chapter 7, 'Discourse' (181–195), rounds out the discussion by tackling the issue of pragmatics. Gao first distinguishes between classroom and real situation Chinese, and then moves to regional accents and dialects, gender, formal and informal style, written versus spoken style, ellipsis and pronoun dropping, conversational fillers, and phraseology. This information will prove beneficial to undergraduate students who are ready to begin language study within China because it points out aspects of the spoken language that tend to frustrate learners emerging from the protected and oversimplified cocoon of a college classroom.

This text includes several helpful features for both students and teachers. Gao includes study questions at the end of each chapter that students will find helpful for learning the material within the preceding text, and that teachers will find helpful for either testing the material or for spawning a discussion of the text. Each chapter also contains a bibliography at the end entitled 'Reference Reading' that provides direction for those desiring to study the chapter's topics in greater detail.

Appendix I contains a comparison between *pinyin* and the Wade–Giles system of Romanization, which should prove helpful for linguists attempting to compare language sources from an earlier period. Perhaps Gao could improve Appendix I by placing the differences between the Wade–Giles system and *pinyin* side–by–side in a more overt manner, rather than forcing the reader to take time to study his separate charts in detail. Appendix II contains a glossary of grammatical terms, which first–year language students and non–linguists will find very helpful for understanding the grammar of their own language in order to understand Chinese grammar by analogy.

One weakness of this text is that all of the examples occur only in *pinyin*. Either traditional or simplified characters with the examples would serve as a pedagogical device for giving first–year students a meaningful encounter with *hanzi* in order to provide one more opportunity for learning the writing system.

One must realize the limitations of *Mandarin Chinese: An introduction*. Linguists who are looking for specific examples of a grammatical phenomenon will find Gao's grammatical description too general and the number of examples too small, and therefore should proceed to a more advanced work that provides greater grammatical precision and more examples, such as Charles N. Li and Sandra A. Thompson's *Mandarin Chinese: A functional reference grammar*. The author states from the beginning, however, that he intends this book to serve as an auxiliary text for undergraduate students, and therefore this book will orient those students who are starting from scratch with helpful background material and

supplementary reading in addition to their main textbook. With regard to Gao's stated purpose, *Mandarin Chinese: An introduction* fulfills its intended function.

Terry Wardlaw, 1593 60th Ave., Albany MN 56307; E-mail address: Terry_Wardlaw@sil.org

Whales, candlelight, and *stuff* like that: General extenders in English discourse. By MARYANN OVERSTREET. New York, Oxford: Oxford University Press. 1999. 171 pp. 2000. Cloth $39.95.

Reviewed by RICHARD L. WATSON
SIL—Graduate Institute of Applied Linguistics

I was looking for a book on literary discourse, so I was at first disappointed with this book. I was also prejudiced that general extenders, such as, 'and stuff (like that)', 'and all (that crap)', 'and everything', 'and blah blah blah', 'or something', especially when peppered throughout a person's speech, were just a lazy, vague means of completing lists, and should be used more sparingly, as the more formal et cetera. This prejudice was reinforced by the frequent use of four letter words in many of Overstreet's examples.

However, prejudices aside, I discovered that this is not only a good example of thorough discourse analysis of an interesting aspect of casual American conversational English, but also of seeing language as a socio–cultural construct. O quotes and cites a great many more noted linguists and relevant work in the area than I had imagined. She first investigates a number of previous studies that analyze general extenders as fulfilling an ideational function, such as being list completers, as occupying the third 'slot' of a list, and as inferring additional or alternate instances of the category the speaker has in mind. O goes into a thorough discussion of categories as studied by such people as Rosch, Channell, Barsalou, Lakoff, Neisser, and Murphy. But even when taking a prototype view of categories, the conclusion of the investigation is that while a speaker may have additional or alternate instances in mind when using general extenders, he/she is more likely to be only indicating the potential existence of other instances and neither the speaker nor hearer may be able to name any. An ideational function may have more or less validity, but it is probably not the relevant concern of speaker or hearer, whereas an interpersonal function is.

Of primary importance under interpersonal function is intersubjectivity and the notion advanced by Schutz (Schutz and Luckmann 1984:77) that people involved in coordinated actions will assume shared knowledge and seek to

sustain it by maintaining the 'reciprocity of perspectives'. Overstreet says (p. 68):

> Given this general operational basis for social interaction, it would not be surprising if interactive talk contained expressions that were conventionally used to invoke an assumption of shared experience or intersubjectivity. One of the most easily recognizable of such expressions is 'you know what I mean', or simply 'you know'. I would like to suggest that general extenders are used to fulfill a similar broad function. In using a general extender, the speaker communicates the following message to the hearer: 'Because we share the same knowledge, experience, and conceptual schemes, I do not need to be explicit; you will be able to supply whatever unstated understandings are required to make sense of the utterance.' Functioning in this way, general extenders are recognizable indications that all talk is, in some sense, incomplete (see Garfinkel 1967), and that each of us expects our interactive partners to collaborate in creating whatever sense of completeness is sufficient for a particular occasion.

Following an example in conversation, O further says:

> In using a general extender, the speaker conveys to the hearer an assumption of shared knowledge, and she invites the hearer to supply whatever unstated understandings are required to make sense of the utterance. By demonstrating an assumption of shared knowledge with the hearer in this manner, the speaker underscores a similarity between the participants. This can be seen either to reaffirm existing familiarity, or to represent a bid to decrease social distance. In turn, by treating the speaker's utterance as unproblematic (and disregarding personal differences in how each assigns meaning to everyday activities), the hearer reciprocally underscores the participants' similarity. General extenders do not implicate categories as much as they implicate intersubjectivity.

Additional interpersonal functions of general extenders are, like sentence adverbs, to express a speaker's attitude towards a message. This may come across as surprise, evaluation, or intensification. For example, in the statement, 'I just found out he ripped me off and everything,' 'and everything' intensifies the victim's shock and the extent to which she felt ripped off.

At this point O begins to subdivide the general extenders by their uses. For example, 'and everything' (an adjunctive extender) indicates a maximum extreme; whereas, 'or anything' (a disjunctive extender) indicates a minimum extreme. An example of the latter is found in a cartoon of two crocodiles lying on a river bank. After dining on the contents of a canoe, one crocodile says to its mate, 'That was incredible. No fur, claws, horns, antlers, or nothin' ... just soft and pink.' A formulaic example of intensification is 'Is this good stuff or what?'

BEST COPY AVAILABLE

O takes off from Gumperz (1984) to discuss the use of general extenders in politeness, that is to try to reduce a sense of imposition—sometimes positively by reinforcing solidarity, sometimes negatively by maintaining independence, e.g. by hedging, 'I'd like to move out there and take over your apartment and all the stuff that's in it an' just buy it from you, or whatever, an' maybe like buy your truck or something. I dunno.'

Another function is as a hedge on certainty or accuracy, e.g., 'They must've broke up or something.' 'An' every year ya know (we) pick ten pounds or somethin'.'

There are many more nuances and examples, and all that stuff, but the important thing is to recognize their interpersonal nature as being more relevant than any ideational significance. The conclusion includes a summary of the typical uses of a number of general extenders. She also includes a number of helpful notes, a considerable bibliography, and an index.

One matter of formatting I found confusing is that all section heads are centered, as if parallel in spite of considerable overlap. Sometimes subordinate headings are italicized, but that is not easy to see. I would have also liked to see some reference to register, other than just the fact that this is 'casual' American speech. Regarding 'and stuff', O relates it to 'and that' in British English, and says (p. 102),

> According to Macaulay (1985:115), the function of and that in these excerpts is 'clearly not set–marking'. Instead, and that is used to appeal to a sense of shared knowledge and functions 'as a kind of punctuation feature, almost the oral equivalent of a comma or a full stop, depending on intonation'.

O then gives an example in which 'and stuff' is repeated ad nauseum (to me), and says (p. 103),

> Just as y'know has become an almost reflex marker for 'I don't have to tell you everything because we share knowledge' (positive politeness), so too is and stuff developing a similar function for some speakers.

I would think that the notion of 'reflex markers' should get more attention, as this is part of what is offensive in some social and literary contexts, as well as to older generations.

It is clear that O has made a large contribution to our understanding of the uses of general extenders and has demonstrated the importance of investigating the interpersonal function of language, especially in areas

where ideational analysis is only partially successful. It is well worth reading.

REFERENCES

Garfinkel, Harold. 1967. Studies in ethnomethodology. Englewood Cliffs, NJ: Prentice-Hall.

Gumperz, John. 1984. Communicative competence revisited. In Meaning, Form and Use in Context: Linguistic Applications, edited by D. Schiffrin, pp. 278–289. Washington, DC: Georgetown University Press.

Macaulay, Ronald. 1985. The narrative skills of a Scottish coal miner. In Focus On: Scotland, edited by Manfred Görlach, pp. 101–124. Amsterdam: John Benjamins.

Schutz, Alfred and Thomas Luckmann. 1984. Strukturen der Lebenswelt, Vol. 2. Frankfurt am Main: Suhrkamp.

Richard L. Watson, 7500 W. Camp Wisdom Rd., Dallas, TX 75236.
E-mail: Dick_Watson@sil.org

■

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

☑ This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

☐ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").